

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 9, 2005, 06:24:33 ; Search time 15452 Seconds
(without alignments)
11499.177 Million cell updates/sec

Title: US-10-600-070-3

Perfect score: 3667
Sequence: 1 tgcctcgcataagagagat.....ctatacataagagctacaa 3667

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:
1: gb Da:
2: gb Hcg:
3: gb In:
4: gb Om:
5: gb Ov:
6: gb Pat:
7: gb Ph:
8: gb Pl:
9: gb Pr:
10: gb Ro:
11: gb Sts:
12: gb Sy:
13: gb Un:
14: gb Vi:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3664	99.9	3668	8	AY221468
2	3660.8	99.8	3664	8	AY221467
3	3659	99.8	85791	8	AB016888
4	1931	52.7	2637	8	AY091075
5	1752.8	47.8	2438	8	AY221469
6	1744.6	47.6	2436	8	AY150490
7	317.6	8.7	150462	8	AP004885
8	304.4	8.3	341	8	AY199896
9	57.6	1.6	7218	6	166494
10	54	1.5	1141	6	AK083744
11	54	1.5	250029	3	AE014820
12	53.8	1.5	205796	2	AL713891
13	53.8	1.5	215524	10	AL603913
14	53.8	1.5	250078	3	AE014829
15	53.8	1.5	333321	3	AC116986
16	53.4	1.5	53.91	8	AY199897
17	53.2	1.5	198573	2	AC110118
18	53	1.4	184039	10	AC139376
19	53	1.4	233373	2	AC134805

20	52.6	1.4	222859	2	AC119799	AC119799 Pan trogl
21	52	1.4	177010	2	AC140372	AC140372 Mus muscu
22	52	1.4	245924	2	BX004773	BX004773 Zebrafish
23	52	1.4	255036	10	AC118206	AC118206 Mus muscu
24	52	1.4	268597	2	AC094787	AC094787 Rattus no
25	51.6	1.4	2000	6	AK655393	AK655393 Sequence
26	51.6	1.4	412213	2	AL590436	AL590436 Human DNA
27	51.6	1.4	110000	2	AC108609	Continuation (3 of
28	51.6	1.4	292533	2	AC107260	AC107260 Rattus no
29	51.4	1.4	139995	10	AC11797	AC11797 Mus muscu
30	51.4	1.4	150579	10	AC126937	AC126937 Mus muscu
31	51.4	1.4	171307	9	AC104020	AC104020 Homo sapi
32	51.4	1.4	193634	2	AC068495	AC068495 Mus muscu
33	51.4	1.4	253094	2	AC113774	AC113774 Rattus no
34	51.2	1.4	227202	2	AC125636	AC125636 Rattus no
35	51.2	1.4	253772	2	AC109777	AC109777 Rattus no
36	51.2	1.4	313096	2	AC131879	AC131879 Rattus no
37	51.2	1.4	339033	2	AC106969	AC106969 Rattus no
38	51	1.4	202358	10	AL596187	AL596187 Mouse DNA
39	51	1.4	269375	2	AC114367	AC114367 Rattus no
40	50.8	1.4	110000	2	AC120236	AC120236 Rattus no
41	50.8	1.4	196519	2	AC127834	AC127834 Rattus no
42	50.8	1.4	266661	2	AC126523	AC126523 Rattus no
43	50.8	1.4	316828	2	AC135386	AC135386 Rattus no
44	50.6	1.4	106693	8	AP004188	AP004188 Oryza sat
45	50.6	1.4	144652	8	AC132003	AC132003 Oryza sat

ALIGNMENTS

RESULT 1	AY221468	3668 bp	DNA	linear	PLN 05-AUG-2003
LOCUS	AY221468				
DEFINITION	Arabidopsis thaliana division protein (ARCE) gene, complete cds;				
ACCESSION	AY221468				
VERSION	AY221468.1	GI:33436338			
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	1 (bases 1 to 3668)				
AUTHORS	OsterYoung,K.W.				
TITLE	ARCE is a J-Domain Plastid Division Protein and an Evolutionary				
JOURNAL	Descendant of the Cyanobacterial Cell Division Protein Ftn2				
PUBMED	Plant Cell 15 (8), 1918-1933 (2003)				
REFERENCE	2 (bases 1 to 3668)				
AUTHORS	Viltha,S., Koksharova,O., van Erp,H., Froehlich,J.E. and				
TITLE	OsterYoung,K.W.				
JOURNAL	Submitted (17-JUN-2003) Department of Plant Biology, Michigan State				
UNIVERSITY	University, 166 Plant Biology Building, East Lansing, MI 48824, USA				
FEATURES	Location/Qualifiers				
source	1..3668				
gene	/organism="Arabidopsis thaliana"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:3702"				
	/ecctype="Wassiljevskija"				
	478..3515				
	/gene="ARC6"				
	/note="synonym: At5g42480"				
	join(<478..984,1067..1291,1379..1639,1723..2710,				
	3000..3133,3225..3515)				
	/gene="ARC6"				
	/product="division protein"				
	join(478..984,1067..1291,1379..1639,1723..2710,3000..3133,				
	3225..3515)				
	/gene="ARC6"				
CDS					

/note="inner envelope membrane-localized; similar to
 cyanobacterial cell division protein sl10169 of
 Synchocystis; contains J-domain"
 /codon_start=1
 /product="division protein"
 /protein_id="AAQ18645.1"
 /db_xref="GI:33436339"
 /translation="MEALSHVGIQLSPOLCRLPATTKLRSHSTTTI CSASKMWAD
 RLIDFNFTSDSSSPATATTAATLVLSIDPSPIDRPERHAPIIDFQVGAQTHPT
 DLRBRAPEARVSKPQPGFDALISRLOLAACETLSNPSRSREBQVGLDEBAT
 VITVDPMKVPALCVTQEGREIYRNBAIKRPLRSPFQDVLTWALAFLDYS
 RDMALDPPDPTGYEFEBEALKILQREGASSLAPDLRAQIDLETLEITTRYLELLG
 LPLGDDYAAARLNLGSGVRLIWSVGGGASALVGLTRKFMNEALRTAARQVUL
 FVATPSNIPAPSEFEVVALALVAQAFI GKPHILQDPAQFOQLQAKWAMEIPAM
 LYDTRNWEIDFGLERGLCALLIGKVDCEMWLGLDESDQYRPAIVAEFLSNSND
 DNDLDELCKLLETWLAGVVPDRDTRDKKPKLDYDDPMTLSYERPEVQSGPL
 AAATMARICAEHYKASAMQALQKPPSPRTDRSAPRKQVQSVSDPQNNVGP
 GEHGVFLAEVRPSENFTNDYAIRGVSSSVDEITVENSAPVDMLKASVKLIAGV
 AIGLISLPSQKYLKSSSFQKQDWSMSDVATIGSVRADSEALPRMDARTAEI
 VSKWOKI KSLAFQPDHRIEMLPVILDRMLKI WDRAAETRAQJLQVVDYTLKLSVDS
 VTVSADOTRALVEATLETESACLSDLVHPENNATDVRTYTRVEMSKGMLTEGSV
 LAS"

ORIGIN

Query Match 99.9%; Score 3664; DB 8; Length 3668;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCTGCATTAGAGAAATACAAATTATAGCAATTTGTCTTGATTTTCAACAAGATTTCCTT 63
 1 TCTGCATTAGAGAAATACAAATTATAGCAATTTGTCTTGATTTTCAACAAGATTTCCTT 60
 DB 64 GGCATAGATTCAATGGCTCTGTTTCTTTTCAATTTACATGTCATATAGTTTCCAAAT 123
 61 GGCATAGATTCAATGGCTCTGTTTCTTTTCAATTTACATGTCATATAGTTTCCAAAT 120
 QY 124 TTACACATTTCAAGTTGATGTTTAAAGAAAGAGAGAAATTGATGGGCTTTTGTGGCTTT 183
 121 TTACACATTTCAAGTTGATGTTTAAAGAAAGAGAGAAATTGATGGGCTTTTGTGGCTTT 180
 DB 184 AAATTTAAAGTGTCAAGAAATTAAGTCATTTGTTTCTGTGCTCTAATNGTGAAT 243
 181 AAATTTAAAGTGTCAAGAAATTAAGTCATTTGTTTCTGTGCTCTAATNGTGAAT 240
 QY 244 GAAGGCAACTCCAAAGCTTTCTAGTGAATATAGTATTTAGACATTTAAACATCATAAA 303
 241 GAAGGCAACTCCAAAGCTTTCTAGTGAATATAGTATTTAGACATTTAAACATCATAAA 300
 DB 304 GTCGTCGCGACGTGTAAACATCATATGATTTTATTTTTCATGATGCTGTGATGTT 363
 301 GTCGTCGCGACGTGTAAACATCATATGATTTTATTTTTCATGATGCTGTGATGTT 360
 QY 364 TTTTGAATTTAACTAATCTAATCAAAATTTCAATTAACCTTGAAGACCAAAACAG 423
 361 TTTTGAATTTAACTAATCTAATCAAAATTTCAATTAACCTTGAAGACCAAAACAG 420
 DB 424 TCTCTTAATATGTAACAAGCAAAAGTTTGTAGTAGGCTTAAAGACACTCCCATG 483
 421 TCTCTTAATATGTAACAAGCAAAAGTTTGTAGTAGGCTTAAAGACACTCCCATG 480
 QY 484 GAAGCTCTGAATCAAGTCGCGCATTTGTCCTCCCATTTCAATTAATGACCAATTAACA 543
 481 GAAGCTCTGAATCAAGTCGCGCATTTGTCCTCCCATTTCAATTAATGACCAATTAACA 540
 DB 544 GCGACGCAAAAGCTCCGACGTAAGCAACAACCTCTACAACTATCTCTCCGACAGAAA 603
 541 GCGACGCAAAAGCTCCGACGTAAGCAACAACCTCTACAACTATCTCTCCGACAGAAA 600
 QY 604 TGAGGCGAAGCGTCTTCTCTCGACCTTCAATTTCACTCCGATTCCTCTCTCTCTCTTC 663
 601 TGAGGCGAAGCGTCTTCTCTCGACCTTCAATTTCACTCCGATTCCTCTCTCTCTCTTC 660
 DB 664 GCCACCGCAACACACCGCACTCTGCTCTCTGCAACATCTATGATGTCGCCGAA 723

DB 661 GCCACCGCAACACACCGCACTCTGCTCTCTGCAACATCTATGATGTCGCCGAA 720
 QY 724 CGCAGCTCCCAATCCCAATTTGATTTCTACAGATATAGAGCTCAACACATTTCTTA 783
 721 CGCAGCTCCCAATCCCAATTTGATTTCTACAGATATAGAGCTCAACACATTTCTTA 780
 QY 784 ACCGATGAATCAAGAAAGCAATTCGAAGCTAGGGTTTCGAAACCGCGCAATTCGGTTTC 843
 781 ACCGATGAATCAAGAAAGCAATTCGAAGCTAGGGTTTCGAAACCGCGCAATTCGGTTTC 840
 DB 844 AGGACGAGCGCTTTAATCAAGCCGGAACAGATTTCTCAAGCTCTTCCGAATCTGTCT 903
 841 AGGACGAGCGCTTTAATCAAGCCGGAACAGATTTCTCAAGCTCTTCCGAATCTGTCT 900
 QY 904 AATCTCTCGCTCTAGAGAGAGTCAATGAAGTCTTCTTGATGATGAAGAGCTACAGTC 963
 901 AATCTCTCGCTCTAGAGAGAGTCAATGAAGTCTTCTTGATGATGAAGAGCTACAGTC 960
 DB 964 ATCACTGATGTTCTTGGATTAAGGTAAATTTGATTTCCGAATTAATGAATTTCTTCGTT 1023
 961 ATCACTGATGTTCTTGGATTAAGGTAAATTTGATTTCCGAATTAATGAATTTCTTCGTT 1020
 QY 1024 TTAATTTCAATGATTTGATTAAGAAAGAACTTTTATCTATGTAAGGTTCTCGGCGCTC 1083
 1021 TTAATTTCAATGATTTGATTAAGAAAGAACTTTTATCTATGTAAGGTTCTCGGCGCTC 1080
 QY 1084 CTGTGATTTGCAAGAAAGTGTGAGACTGAGATAGTTCTTCCGGTTGTGAGGCTCTGCT 1143
 1081 CTGTGATTTGCAAGAAAGTGTGAGACTGAGATAGTTCTTCCGGTTGTGAGGCTCTGCT 1140
 DB 1144 TAAAGAAAGGTCCTTAAGTGTGTTTAAAGCAAGTGTGTTTAAAGTGTGCGCTTCGCTT 1203
 1141 TAAAGAAAGGTCCTTAAGTGTGTTTAAAGCAAGTGTGTTTAAAGTGTGCGCTTCGCTT 1200
 QY 1204 TCTGCATGTCGAGAGGACCTTAAGTGTGATTTTCAACCTGATTTTATTAATGATTAAGA 1263
 1201 TCTGCATGTCGAGAGGACCTTAAGTGTGATTTTCAACCTGATTTTATTAATGATTAAGA 1260
 DB 1264 GTTTGTTGAGAGACTTTTGAAGCTTTTACAGATGTTTGAAGCTTTGTTGAATTTTGAACG 1323
 1261 GTTTGTTGAGAGACTTTTGAAGCTTTTACAGATGTTTGAAGCTTTGTTGAATTTTGAACG 1320
 QY 1324 AGCGTGTGCTTTAATGAAGCTTTTCTGATTTGATTAATCTTGTATTAAGCTTTGTGTAAGA 1383
 1321 AGCGTGTGCTTTAATGAAGCTTTTCTGATTTGATTAATCTTGTATTAAGCTTTGTGTAAGA 1380
 DB 1384 GGAAGAGCAAGTAAAGCTTGGACACCGGATTTACGTGCAAAATTGATGAAGAATTGAGA 1443
 1381 GGAAGAGCAAGTAAAGCTTGGACACCGGATTTACGTGCAAAATTGATGAAGAATTGAGA 1440
 QY 1444 GATCACTCCGCTTATGTCCTTGGAGCTACTTGGCTTACCGCTTGGTATGATTAACGCTGC 1503
 1441 GATCACTCCGCTTATGTCCTTGGAGCTACTTGGCTTACCGCTTGGTATGATTAACGCTGC 1500
 DB 1504 GAAAGAATTAATGTTTAAAGCGGTGTCGGAATATTTTGTGTCCTGTGAGAGAGGTG 1563
 1501 GAAAGAATTAATGTTTAAAGCGGTGTCGGAATATTTTGTGTCCTGTGAGAGAGGTG 1560
 QY 1564 AGCATCAGCTCTTGTGGGGGTTTGAACCGGTGAAGATTTAATGAAGAGCGTTTTTACG 1623
 1561 AGCATCAGCTCTTGTGGGGGTTTGAACCGGTGAAGATTTAATGAAGAGCGTTTTTACG 1620
 DB 1624 AATGACAGTGTGAGAGGATTAAGTATTAAGTATTAATCTTTTAAATTTCTTAAAGCATGA 1683
 1621 AATGACAGTGTGAGAGGATTAAGTATTAAGTATTAATCTTTTAAATTTCTTAAAGCATGA 1680
 QY 1684 TATTAATTAAGTGTTCATTTTAATGATATGTTGTGTGATGATCTTTTGTAGCT 1743
 1681 TATTAATTAAGTGTTCATTTTAATGATATGTTGTGTGATGATCTTTTGTAGCT 1740
 DB 1744 ACCCAAGCAATATTCAGAGAGTCAATTTGAAGTTTACGAATTTGACATTTGCTCTGTG 1803

Db 1741 ACCCCAGCATATATCCAGCAGAGTCATTTGAGTTTACGAAAGTTCACCTTCTGTG 1800
 Qy 1804 GCTCAAGCTTTTATGTTAGAGAGCCACCTTTTACAGATGCGATTAAGCAATTCAG 1863
 Db 1801 GCTCAAGCTTTTATGTTAGAGAGCCACCTTTTACAGATGCGATTAAGCAATTCAG 1860
 Qy 1864 CAATTCAGCAGGCTTAAGTAAAGCTATAGAGATTCCTGCGATGTTGATATACACG 1923
 Db 1861 CAATTCAGCAGGCTTAAGTAAAGCTATAGAGATTCCTGCGATGTTGATATACACG 1920
 Qy 1924 AATTAATGGAGATAGACTTCGTCTAGAAAGGGAGCTCTGCACTGCTTATAGCCAA 1983
 Db 1921 AATTAATGGAGATAGACTTCGTCTAGAAAGGGAGCTCTGCACTGCTTATAGCCAA 1980
 Qy 1984 GTTGAATAGCCGATAGTGTGGGCTTACAGATGAGATTCATATAGGAATCA 2043
 Db 1981 GTTGAATAGCCGATAGTGTGGGCTTACAGATGAGATTCATATAGGAATCA 2040
 Qy 2044 GCTATTGTGAGTTGTTTGGAGAAATTCAAATGCTGATGCAATGATCTCCCTGA 2103
 Db 2041 GCTATTGTGAGTTGTTTGGAGAAATTCAAATGCTGATGCAATGATCTCCCTGA 2100
 Qy 2104 CTATGCAAAATTTGTGAAACCTGCTGGCAGAGGGTGTCTTCTAGGTTCAAGACA 2163
 Db 2101 CTATGCAAAATTTGTGAAACCTGCTGGCAGAGGGTGTCTTCTAGGTTCAAGACA 2160
 Qy 2164 AAAGATTAATAATTTAACTCGGGGACCTACATAGATTCCTATAGTTTGAATACCTG 2223
 Db 2161 AAAGATTAATAATTTAACTCGGGGACCTACATAGATTCCTATAGTTTGAATACCTG 2220
 Qy 2224 GAAAGATGAGAGTTCAGAGGTTCTCTTACCTGCTGCAACTATGCGAAGAT 2283
 Db 2221 GAAAGATGAGAGTTCAGAGGTTCTCTTACCTGCTGCAACTATGCGAAGAT 2280
 Qy 2284 GAGCCGAGCATGTGAAGCTAGTGTATAGCAGCAGCTGCAAGAAATTTCTCCGCG 2343
 Db 2281 GAGCCGAGCATGTGAAGCTAGTGTATAGCAGCAGCTGCAAGAAATTTCTCCGCG 2340
 Qy 2344 TATACATATGAATCTGGGCTGAACCCAGAGATGTCAGAGAGATTTAGTATGAT 2403
 Db 2341 TATACATATGAATCTGGGCTGAACCCAGAGATGTCAGAGAGATTTAGTATGAT 2400
 Qy 2404 CTTGTGTGTAACTATAGCCGATGATGAGCTGCTGCTTATTTGCAAGAGCTGA 2463
 Db 2401 CTTGTGTGTAACTATAGCCGATGATGAGCTGCTGCTTATTTGCAAGAGCTGA 2460
 Qy 2464 AGACCTCTGAAAACTTTGAAACTATATGATTAATTCAGAGCTGGGGTCTCAGAGAGT 2523
 Db 2461 AGACCTCTGAAAACTTTGAAACTATATGATTAATTCAGAGCTGGGGTCTCAGAGAGT 2520
 Qy 2524 AGCCTTATGAACTACTCTGTGAAATGTCGCTGCTGATATGTTAAAGAGGCAAGTGTG 2583
 Db 2521 AGCCTTATGAACTACTCTGTGAAATGTCGCTGCTGATATGTTAAAGAGGCAAGTGTG 2580
 Qy 2584 AAGATCTAGCTGCTGATGAGCAATTCAGATTCATCTGCTCAGGCAAGATTTT 2643
 Db 2581 AAGATCTAGCTGCTGATGAGCAATTCAGATTCATCTGCTCAGGCAAGATTTT 2640
 Qy 2644 CTTAAAGCAGCTCATCTTTCAACGCAAGATATGTTTCTTATAGAAATCTGATGTC 2703
 Db 2641 CTTAAAGCAGCTCATCTTTCAACGCAAGATATGTTTCTTATAGAAATCTGATGTC 2700
 Qy 2704 GCTACCAATAGTATGATTAATGATGCAATTTCTATATCTGCAATGCTCAAAATATGC 2763
 Db 2701 GCTACCAATAGTATGATTAATGATGCAATTTCTATATCTGCAATGCTCAAAATATGC 2760
 Qy 2764 TTTGTTTGTAGCTAAGAACATATGTCCTCACTTATATACATGTCCTCAAAAGTTGATCAAG 2823
 Db 2761 TTTGTTTGTAGCTAAGAACATATGTCCTCACTTATATACATGTCCTCAAAAGTTGATCAAG 2820
 Qy 2824 ATTAAACAAGTTGCTGAGTAAATTTCACTAATATATGCTGCTGAAATTTTGTATCAAACTG 2883
 Db 2821 ATTAAACAAGTTGCTGAGTAAATTTCACTAATATATGCTGCTGAAATTTTGTATCAAACTG 2880

Qy 2884 TAGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTTAAATTAACGTAGATTTAGAGA 2943
 Db 2881 TAGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTTAAATTAACGTAGATTTAGAGA 2940
 Qy 2944 TTGCTTATGATGCTGCTTGTTCAACTTTTCTTCTGATTTTCTTTTTCGATTTAGG 3003
 Db 2941 TTGCTTATGATGCTGCTTGTTCCACTTTTCTTCTGATTTTCTTTTTCGATTTAGG 3000
 Qy 3004 GTCAGCAGAGCTGACGATTCAGAGACCTTCCAGATGAGATGCTGAGACGAGAGA 3063
 Db 3001 GTCAGCAGAGCTGACGATTCAGAGACCTTCCAGATGAGATGCTGAGACGAGAGA 3060
 Qy 3064 TATAGTATCCAGAGTGGCAGAAATTAAGTCTCTGCTTTTGGGCTGATCAACCCATAGA 3123
 Db 3061 TATAGTATCCAGAGTGGCAGAAATTAAGTCTCTGCTTTTGGGCTGATCAACCCATAGA 3120
 Qy 3124 AATGTTACAGAGGTGAGGAAATTAATCTACATTCATCAATGTTGAAACTGTTGG 3183
 Db 3121 AATGTTACAGAGGTGAGGAAATTAATCTACATTCATCAATGTTGAAACTGTTGG 3180
 Qy 3184 ACATGATTAATAGTCTGCTGCTTGTGATTCCTGTTTATTAATAGTTTGAATGGGCGAA 3243
 Db 3181 ACATGATTAATAGTCTGCTGCTTGTGATTCCTGTTTATTAATAGTTTGAATGGGCGAA 3240
 Qy 3244 TGTGAAGATTTTGAAGTCAAGACAGCTGAACTGCGCAGCTTGGGTTTATGAT 3303
 Db 3241 TGTGAAGATTTTGAAGTCAAGACAGCTGAACTGCGCAGCTTGGGTTTATGAT 3300
 Qy 3304 ATACACTGTTGAAACTATCTGTTGACAGATGACAGTCTCAGCAGATGAAACCCGTGCTC 3363
 Db 3301 ATACACTGTTGAAACTATCTGTTGACAGATGACAGTCTCAGCAGATGAAACCCGTGCTC 3360
 Qy 3364 TGTGGAAGCAACTCTGGAAGAGTCTGCTGCTATCTGATTTGATTCATCCAGAAACA 3423
 Db 3361 TGTGGAAGCAACTCTGGAAGAGTCTGCTGCTATCTGATTTGATTCATCCAGAAACA 3420
 Qy 3424 ATGCTACTGATGTCAGAACCTACACAACAAGATACGAATTTTCTGCTCAAGTCAAGGT 3483
 Db 3421 ATGCTACTGATGTCAGAACCTACACAACAAGATACGAATTTTCTGCTCAAGTCAAGGT 3480
 Qy 3484 GGAATATCACTGAAGGCTCTGTTCTTGCATCAATTAATCTCATATGATGATGCTGAG 3543
 Db 3481 GGAATATCACTGAAGGCTCTGTTCTTGCATCAATTAATCTCATATGATGATGCTGAG 3540
 Qy 3544 CTTGCGAGATTTCTTTTGTCTGTTAAATCTCTCTCAAGTATGTTTAAATGAAACA 3603
 Db 3541 CTTGCGAGATTTCTTTTGTCTGTTAAATCTCTCTCAAGTATGTTTAAATGAAACA 3600
 Qy 3604 CAAAAAATTAAGCTTCTGTCGACACCCCTTTCTGATCTAACTAATCAATTAAGGCT 3663
 Db 3601 CAAAAAATTAAGCTTCTGTCGACACCCCTTTCTGATCTAACTAATTAAGGCT 3660
 Qy 3664 ACNA 3667
 Db 3661 ACNA 3664

RESULT 2
 AY221467
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 3664)
 REFERENCE
 Vitha,S., Froehlich,J.E., Koksharova,O., Pyke,K.A., Van Eyrp,H. and

Oteyyoung, K.W.
Atc6 is a J-Domain Plastid Division Protein and an Evolutionary
Descendant of the Cyanobacterial Cell Division Protein Ftn2
JOURNAL
PUBMED
Plant Cell 15 (8), 1918-1933 (2003)
12897262
2 (bases 1 to 3664)
Vitha, S., Koksharova, O., van Erp, H., Froehlich, J.E. and
Oteyyoung, K.W.
Direct Submission
Submitted (17-JUN-2003) Department of Plant Biology, Michigan State
University, 166 Plant Biology Building, East Lansing, MI 48824, USA
Location/Qualifiers
Source
1..3664
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/ecotype="Ws-2"
/note="genotype: arc6 mutant"
1..3664
/gene="arc6"
/allele="arc6-1"
join(<478..984,1067..1291,1379..>1621)
/gene="arc6"
/allele="arc6-1"
/product="truncated division protein"
join(478..984,1067..1291,1379..1621)
/gene="arc6"
/allele="arc6-1"
/note="truncation caused by premature stop codon"
/codon_start=1
/product="truncated division protein"
/protein_id="AAQ18644.1"
/db_xref="GI:33436275"
/translation="MEALSHVIGLSPPOLCRLPATTKLRSHNSTTICSAKMWAD
RLSDPRTDSSSSPRTATRTTIVSLPSIDRPRHPPIPIIDYQVCAQTHPT
DGRRAEAVSRKPPQFSDALISRQILQACETLSPRSRKRTNGLDDEEAT
VITDVPDKVGA/CVLQEGEETIVRVEBALKEPLPSFKQDVVLAALFADVS
RDMAALDPDITGYEFVEBALKLOEGASLAPDRADIDETLEITRYYVELLIG
LPIDDYAAKRLNGLSGVRNLMISVGGSALVGLTRKFMNEARL"

ORIGIN

Query Match 99.8%; Score 3660.8; DB 8; Length 3664;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3662; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 TCTGCATTAGGGAATTAACAATTATTAAGCAATTTGTCGATTTCAACAAGATTGCTT 63
DB 1 TCTGCATTAGGGAATTAACAATTATTAAGCAATTTGTCGATTTCAACAAGATTGCTT 60
QY 64 GGCATATAGATTCAATGGCTCTGTTGCTTTACATTTACATGTCATATAGTTGCAAT 123
DB 61 GGCATATAGATTCAATGGCTCTGTTGCTTTACATTTACATGTCATATAGTTGCAAT 120
QY 124 TTTAACAATTCAGTTGAGATGTTAAAGAAAGAGGGAATTGATGGGTTTGTGGGTTT 183
DB 121 TTTAACAATTCAGTTGAGATGTTAAAGAAAGAGGGAATTGATGGGTTTGTGGGTTT 180
QY 184 AAACCTTAAGTGTCAAGAAATTAAGTCATTTGTTCTGTGCTCAATATGTTAAAT 243
DB 181 AAACCTTAAGTGTCAAGAAATTAAGTCATTTGTTCTGTGCTCAATATGTTAAAT 240
QY 244 GAAAGCAACTCCAAGGTTCTTAGTGAATAGATTATTTAGACGATTTAATCATATAA 303
DB 241 GAAAGCAACTCCAAGGTTCTTAGTGAATAGATTATTTAGACGATTTAATCATATAA 300
QY 304 GTCCGTGGCGACGTGAACATCAATGATGTTTTTTTATTTTTCAGTACGTGTGATGTT 363
DB 301 GTCCGTGGCGACGTGAACATCAATGATGTTTTTTTATTTTTCAGTACGTGTGATGTT 360
QY 364 TTTTGAATTTAACTTAACTAACTCAAAATTCATAAAACCTTGAAGCAACCAACAG 423
DB 361 TTTTGAATTTAACTTAACTCAAAATTCATAAAACCTTGAAGCAACCAACAG 420

QY 424 TCTCTCAATATGTAAAAACGAACAAAGTTTTGTGATAGCTTAAAGACACTCCATG 483
DB 421 TCTCTCAATATGTAAAAACGAACAAAGTTTTGTGATAGCTTAAAGACACTCCATG 480
QY 484 GAAGCTCTGATGACCGTCGGCATTTGGTCTCTCCCATTTCAATTATAGCGATTACACCG 543
DB 481 GAAGCTCTGATGACCGTCGGCATTTGGTCTCTCCCATTTCAATTATAGCGATTACACCG 540
QY 544 GCGACGACAAAGCTCCGACGTAGCCACACMACCTCTACAACTATCTGCTCCGACGAA 603
DB 541 GCGACGACAAAGCTCCGACGTAGCCACACMACCTCTACAACTATCTGCTCCGACGAA 600
QY 604 TGGGCGGACCGCTCTTCTCTCCGACTTCAATTTCACTCCGATTTCTCTCTCTCTT 663
DB 601 TGGGCGGACCGCTCTTCTCTCCGACTTCAATTTCACTCCGATTTCTCTCTCTCTT 660
QY 664 GCCACGCGCACACACCGGCACTCTGCTCTGCGACACATCTATGATCGCCGAA 723
DB 661 GCCACGCGCACACACCGGCACTCTGCTCTGCGACACATCTATGATCGCCGAA 720
QY 724 CGCCACGTCCCAATCCCATTTGATTTCTACCAAGTATTAAGAGCTCAAAACATTTCTTA 783
DB 721 CGCCACGTCCCAATCCCATTTGATTTCTACCAAGTATTAAGAGCTCAAAACATTTCTTA 780
QY 784 ACCGATGAATCAAGAGACATTCGAAGCTAGGGTTTCGAACCGCGCAATTCGGTTTC 843
DB 781 ACCGATGAATCAAGAGACATTCGAAGCTAGGGTTTCGAACCGCGCAATTCGGTTTC 840
QY 844 AGGACGACGCTTTAATCAACCGGAGACAGATTCTTCAAGCTGCTTGCAGAACTCTGCT 903
DB 841 AGGACGACGCTTTAATCAACCGGAGACAGATTCTTCAAGCTGCTTGCAGAACTCTGCT 900
QY 904 AATCTCGGCTGAAGAGAGTCAATGAAGAGTCTTCTTGATGATGAAGAGCTACAGTC 963
DB 901 AATCTCGGCTGAAGAGAGTCAATGAAGAGTCTTCTTGATGATGAAGAGCTACAGTC 960
QY 964 ATCACTGATGTTCTTGGATTAAGTATTCGATTTCCGAATTAATTAAGTTTCTTCTGTT 1023
DB 961 ATCACTGATGTTCTTGGATTAAGTATTCGATTTCCGAATTAATTAAGTTTCTTCTGTT 1020
QY 1024 TTAAATTTCAATATTAAGTAAAGAAAGAACTTTTATCTATGTAAGGTTCTGCGGCTCT 1083
DB 1021 TTAAATTTCAATATTAAGTAAAGAAAGAACTTTTATCTATGTAAGGTTCTGCGGCTCT 1080
QY 1084 CTGTGATATTCAGAAAGGTGTGAGACTGAGATAGTTCTTCCGTTGTGTAGGCTCTGCT 1143
DB 1081 CTGTGATATTCAGAAAGGTGTGAGACTGAGATAGTTCTTCCGTTGTGTAGGCTCTGCT 1140
QY 1144 TAAAGAGAGGTTGCTTAAGTCTGTTTAAAGCAAGTGTGTTTAAAGTGTGCGTCTGCTT 1203
DB 1141 TAAAGAGAGGTTGCTTAAGTCTGTTTAAAGCAAGTGTGTTTAAAGTGTGCGTCTGCTT 1200
QY 1204 TCTCGATGTCGAGAGGAGTGTATGTCATTCGATTCGATTTTATTACTGTTATGA 1263
DB 1201 TCTCGATGTCGAGAGGAGTGTATGTCATTCGATTCGATTTTATTACTGTTATGA 1260
QY 1264 GTTGTGTTAGAGAGCTTTGAAGCTTTTACAGATGTTTACCTTGTGTTAATTTGACG 1323
DB 1261 GTTGTGTTAGAGAGCTTTGAAGCTTTTACAGATGTTTACCTTGTGTTAATTTGACG 1320
QY 1324 AGCGTGTGCTTTAAGAACTTTCTTGATTTGATCTTTGTTATGATCTTGTGAGGA 1383
DB 1321 AGCGTGTGCTTTAAGAACTTTCTTGATTTGATCTTTGTTATGATCTTGTGAGGA 1380
QY 1384 GGAAGAGCAAGTATGCTTTCACACCGGATTTAGTGACAAATTTGATGAGACTTTGGAAGA 1443
DB 1381 GGAAGAGCAAGTATGCTTTCACACCGGATTTAGTGACAAATTTGATGAGACTTTGGAAGA 1440
QY 1444 GATCACTCGCGGTTATGTTGTGAGGCTAATTTGAGCTTAACGCTTGTGATGATTAACGCTGC 1503
DB 1441 GATCACTCGCGGTTATGTTGTGAGGCTAATTTGAGCTTAACGCTTGTGATGATTAACGCTGC 1500
QY 1504 GAAAAGCTAAATGTTTAAAGCGGTGTGCGGAATATTTGTGTGTTGTGAGAGAGGTGG 1563

[illegible]

Db	2581	AAGATCTAGCTGCTGGTGTGGCAATTGACATGATTTCACGTTCGAGCAGAAATATTTT	26410
Qy	2644	CTTAAAGACAGCTCATCTTTTCAAACGAGAGATATGTTTCTTCTATGGAATCTGATGTC	27030
Db	2641	CTTAAAGACAGCTCATCTTTTCAAACGAGAGATATGTTTCTTCTATGGAATCTGATGTC	27000
Qy	2704	GCTACCATAGGTATGATTTAAATGATNGCAATTTTATATATCTGCATGCTCAAAATATGTC	27630
Db	2701	GCTACCAATGGTATGATTTAAATGATNGCAATTTTATATATCTGCATGCTCAAAATATGTC	27600
Qy	2764	TTGTTTTGTGAGCTAAGAACATATGTTCCCACTTATATACATGTCCCAAAAGTTGTATCCAAAG	28230
Db	2761	TTGTTTTGTGAGCTAAGAACATATGTTCCCACTTATATATATGTCCCAAAAGTTGTATCCAAAG	28200
Qy	2824	ATTAAACAAGTTGCTGAGTAATAATTTCACTAATTTATGCTGCTGAATTTTTCATCAACTG	28830
Db	2821	ATTAAACAAGTTGCTGAGTAATAATTTCACTAATTTATGCTGCTGAATTTTTCATCAACTG	28800
Qy	2884	TAGACAGAAATGTAATTTTCACTCTCAACATTTCTGTTTAAATATAGTAGAGATTAGAGA	29430
Db	2881	TAGACAGAAATGTAATTTTCACTCTCAACATTTCTGTTTAAATATAGTAGAGATTAGAGA	29400
Qy	2944	TTGCCTTAGTGTGGCTTTGTCCAACTTTTCTTCTGTAATTTTTCCTTTTCTTTTGAATTTAGG	30030
Db	2941	TTGCCTTAGTGTGGCTTTGTCCAACTTTTCTTCTGTAATTTTTCCTTTTCTTTTGAATTTAGG	30000
Qy	3004	GTCAGTCAGAGCTGACGATTCAGAGACACTTCCAGAAATGATGCTAGAGCTGACAGAA	30630
Db	3001	GTCAGTCAGAGCTGACGATTCAGAGACACTTCCAGAAATGATGCTAGAGCTGACAGAA	30600
Qy	3064	TATAGTATCCAGTGTGACGAGAAATTAAGTCTCTGGCTTTTGGGCTGTATCAACCGATATGA	31220
Db	3061	TATAGTATCCAGTGTGACGAGAAATTAAGTCTCTGGCTTTTGGGCTGTATCAACCGATATGA	31200
Qy	3124	AATGTTACAGAGGTGAGGAGAAATTAATCTAATTAATCAATTAATGTGTGAAAACCTGTGG	31830
Db	3121	AATGTTACAGAGGTGAGGAGAAATTAATCTAATTAATCAATTAATGTGTGAAAACCTGTGG	31800
Qy	3184	ACATGATTAATAGTCTGTGCTTGTCTTTGATTTCTGTATTTATAGATTTTGGATGGCGAA	32430
Db	3181	ACATGATTAATAGTCTGTGCTTGTCTTTGATTTCTGTATTTATAGATTTTGGATGGCGAA	32400
Qy	3244	TGCTGAGATTTTGGACTGACAGAGCAGCTGAAATCTGCGACGCTTGGTGTATATGATTT	33030
Db	3241	TGCTGAGATTTTGGACTGACAGAGCAGCTGAAATCTGCGACGCTTGGTGTATATGATTT	33000
Qy	3304	ATACACCTTTGAAACTATCTGTGACAGCTGACAGCTGACAGATGGAACCCGTGCTC	33630
Db	3301	ATACACCTTTGAAACTATCTGTGACAGCTGACAGCTGACAGATGGAACCCGTGCTC	33600
Qy	3364	TGTTGGAAGCAACTGTGAGAGAGTCTGTTGTCTATCTGATTTTGGTTCATCCAGAAAACA	34220
Db	3361	TGTTGGAAGCAACTGTGAGAGAGTCTGTTGTCTATCTGATTTTGGTTCATCCAGAAAACA	34200
Qy	3424	ATGCTACTGATGTGAGAACTTACACACAAGATACGAAGTTTTCGTGTCCAACTCAGGCT	34830
Db	3421	ATGCTACTGATGTGAGAACTTACACACAAGATACGAAGTTTTCGTGTCCAACTCAGGCT	34800
Qy	3484	GGAATAATCACTGAAAGGCTGTGTTCTTGATGATATAATAACATATGAGCATGTCTGAG	35430
Db	3481	GGAATAATCACTGAAAGGCTGTGTTCTTGATGATATAATAACATATGAGCATGTCTGAG	35400
Qy	3544	CTTGGAGATTTCTCTTTGTTCTGTAAATCTCTCTCTAAGTTAGTGTATTAATGAAACA	36030
Db	3541	CTTGGAGATTTCTCTTTGTTCTGTAAATCTCTCTCTAAGTTAGTGTATTAATGAAACA	36000
Qy	3604	CAAAAATTAATACGTTCTTGGCACAACCTTTTCTTATCTTAACTTAACTAATAGGCT	36630
Db	3601	CAAAAATTAATACGTTCTTGGCACAACCTTTTCTTATCTTAACTTAACTAATAGGCT	36600
Qy	3664	ACAA 3667	
Db	3661	ACAA 3664	

RESULT 3
LOCUS AB016888 85791 bp DNA linear PLN 14-FEB-2004
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, pl clone:MDH9.
ACCESSION AB016888 BA000015
VERSION AB016888.1 GI:3449329
KEYWORDS
SOURCE
ORGANISM Arabidopsis thaliana (chale cress)
Arabidopsis thaliana
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta, eudicotyledons, core eudicots;
rosids, euroside II; Brassicales; Brassicaceae; Arabidopsis.
1
Asamizu,E., Sato,S., Kaneko,T., Nakamura,Y., Kotani,H., Miyajima,N.
and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
Sequence features of the regions of 1,081,958 bp covered by
seventeen physically assigned P1 and YAC clones
JOURNAL DNA Res. 5 (6), 379-391 (1998)
MEDLINE 99156233
PubMed 10048488
REFERENCE 2 (basee 1 to 85791)
AUTHORS Nakamura,Y.
TITLE Direct Submission
Submitted (18-AUG-1998) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:yakunam@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Addresses for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MDH9
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://complib.ornl.gov/Grail-1.3/),
GENSCAN (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K5J14 and the 3' clone is K16E1.
Location/Qualifiers
FEATURES
source
1..85791
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MDH9"
/clone_1lb="Mitau1 P1"
/ecotype="Columbia"
complement(59..239)
/note="CDS is reported in Acc# AB023032
gene_id:K5J14.12
unknown protein"
/number=1
/evidence=not_experimental
complement(join(2056..2079,2278..2393,2991..3633))
/note="contains similarity to unknown protein
gene_id:MDH9.1
pit||T04864"
/pseudo
/codon_start=1
/evidence=not_experimental
join(5739..6138,6462..6595)

CDS
/note="unnamed protein product; gene_id:MDH9.2
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10474.1"
/db_xref="GI:9759472"
/translation="MKDEFAEEDRKKSNLSLAKNSVLPFERGRFYEEYAKNE
RLRRKGEEAVGGVKGTPYINLVNHEPMTNRGRGAKESITKTVSVNVESSSIPR
YSLRSMAKENRRKPPPLPNAVAVSAMKSVTRKEERLSLLCTCACTSSGLLAFLRTGS
EPGVHKKRTLPSPALV"
complement(join(6868..7911,7989..8414,8555..8671,
8754..9137))
/note="gene_id:MDH9.3"
/codon_start=1
/evidence=not_experimental
/product="arm repeat containing protein"
/protein_id="BAB10475.1"
/db_xref="GI:9759470"
/translation="MDTERATGDAENRDEBTAEPIHDEVYDAVEIHEEYKEDD
DCGLVSDIVSIVEFLDQINGYRRTQKCEFNVRRLKLIPLDEIRGEFSCKHF
LNLRLKVFLLAAKLTLETCNSGKIYALDETMTTFHSIYEKLNRLVAVAPDEMI
SGAKDEIDSLCKQLKAKKRTDODIELADVMVVSFKDPPNADSAIIEFLAKLE
LQTIIDLTETITAIQSLIQDKGGLNETKOHIIELNKKFKQLGELATDILVGVNPK
AITKTSLLIPHEFLCPIETLEIMLDPVIIATGQYKESIQKPFDGHKTCPTROT
DHLSLAPNRLAKNLIWQMCCKNFKIPKSVSPSQNEQDEVSILVEALSSQLEQ
RRSVKQMRLLARENPNRVLIAVAGAIPLVQLSYPDSGIQNAVTTLLNISIDEN
KRLISNGBAIPNIIIELENGNREARNSAALFSLMLDENKVTIGISNGIPLVLDL
OHTLKGKQALPALFNLISNANKRAIDAGIVQPLNLKDKNIGMIDALSILL
LASHPGROALIQGLSFETLVEFTRGTPNKKCAVSLLELSSNSSFLAALQGV
YRIIVETTSQTRAKRKANALIQLSKSRQT"
10888..12579
/note="unnamed protein product; gb|AA07802.1
gene_id:MDH9.4
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10476.1"
/db_xref="GI:9759471"
/translation="MMISEKPLGEESIRDLLEVLTVSRRLVKSQKLKKIKHTEV
VEBEIVAGVFNCLISVGCRAVDGTGEDPDSNKMSSASEGKGLMTICGEERL
DCSYVVRBEPFKNNRKYLDAGDGYRKHVYLPDILMCLMRLLPLTSLNHLVCK
KMSMANQRFLOMRREGSFQTPWMLPALNCGCSGDIHGVDSQDKNRIETDL
KGFMTSVSIIHEIITVIGSRMDRNSFSHRGILVPSPIKAMRKIASNRHAKSLPI
VGAETVTSFSTWQTQNRQDRFHLRSVGESDVEDEPHRLSVRQNRNSADQNGTK
SHRLIQKLDRLNRNSKRFVLLAIGTGLFDEPLDSGEIYDSATVTSMDRLPMGF
GVVSGCIIINGIFVAYSENDKLSGYIEGFTITQSPILPVEHFVEPLVSCNRL
FMLSVCBQBGQIGRNRKAVKMELDVYITVEVSNHPAPMDMATVYSDNI
LMQIEMPKFIFGYLSFTYCDVUTBEASNRHVSRRNSQCLNLSCTNKITALHLT"
complement(15087..16778)
/note="unnamed protein product; gb|AA07802.1
gene_id:MDH9.5
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB10477.1"
/db_xref="GI:9759472"
/translation="MMISEKPLGEESIRDLLEVLTVSRRLVKSQKLKKIKHTEV
EBDEVVAGVFNCLISVGCRAVDGTGEDPDSNKMSSASEGKGLMTICGEERL
DCSYVVRBEPFKNNRKYLDAGDGYRKHVYLPDILMCLMRLLPLTSLNHLVCK
KMSMANQRFLOMRREGSFQTPWMLPALNCGCSGDIHGVDSQDKNRIETDL
KGFMTSVSIIHEIITVIGSRMDRNSFSHRGILVPSPIKAMRKIASNRHAKSLPI
VGAETVTSFSTWQTQNRQDRFHLRSVGESDVEDEPHRLSVRQNRNSADQNGTK
SHRLIQKLDRLNRNSKRFVLLAIGTGLFDEPLDSGEIYDSATVTSMDRLPMGF
GVVSGCIIINGIFVAYSENDKLSGYIEGFTITQSPILPVEHFVEPLVSCNRL
FMLSVCBQBGQIGRNRKAVKMELDVYITVEVSNHPAPMDMATVYSDNI
LMQIEMPKFIFGYLSFTYCDVUTBEASNRHVSRRNSQCLNLSCTNKITALHLT"
complement(join(18111..18193,18281..18535,18773..19001,
19088..19195,19307..19447,19542..19952,20129..20245))
/note="unnamed protein product; gene_id:MDH9.6
similar to unknown protein
bp|P42251"
/codon_start=1
/evidence=not_experimental

CDS

CDS

CDS

CDS

QY 1261 TGAGTTGTGAGGAACTTTGAAAGCTTTTACAGAGTGTGACCTTGTGTAAATTG 1320
Db 64857 TGAGTTGTGAGGAACTTTGAAAGCTTTTACAGAGTGTGACCTTGTGTAAATTG 64916
QY 1321 ACCAGCCTTGCTTTATTAAGAACTTTTGTGATTTGATTTGATTTGATTTGATTTG 1380
Db 64917 ACCAGCCTTGCTTTATTAAGAACTTTTGTGATTTGATTTGATTTGATTTGATTTG 64976
QY 1381 GAGAGAAAGAGCAAGTACCTTGACCGGATTTACGTGACAAATTTGATGAGACTTTGGA 1440
Db 64977 GAGAGAAAGAGCAAGTACCTTGACCGGATTTACGTGACAAATTTGATGAGACTTTGGA 65036
QY 1441 AGAGATCACTCCGCGTTATGTCTTGAGACTATTTGGCTTACCGCTTGTGTATGACG 1500
Db 65037 AGAGATCACTCCGCGTTATGTCTTGAGACTATTTGGCTTACCGCTTGTGTATGACG 65096
QY 1501 TCGGAAAAAGCTAAATGGTTTAAAGCGGTGCGGAATATTTTGTGTCTGTGAGAGAG 1560
Db 65097 TCGGAAAAAGCTAAATGGTTTAAAGCGGTGCGGAATATTTTGTGTCTGTGAGAGAG 65156
QY 1561 TGAGCATGAGCTCTTGTGGGGGTTTGAACCGTGAAGAGTTTATGATGAGGCGTTT 1620
Db 65157 TGAGCATGAGCTCTTGTGGGGGTTTGAACCGTGAAGAGTTTATGATGAGGCGTTT 65216
QY 1621 ACCGATGACAGCTGCTGAGCAGGTATACGTTTGAATACCTTTTATTTTATTTTGA 1680
Db 65217 ACCGATGACAGCTGCTGAGCAGGTATACGTTTGAATACCTTTTATTTTATTTTGA 65276
QY 1681 TGATATTAATTTAGGTTTCTCATTTTAAATGATTTGTGTGTGTGTGTGTGTGTGT 1740
Db 65277 TGATATTAATTTAGGTTTCTCATTTTAAATGATTTGTGTGTGTGTGTGTGTGTGT 65336
QY 1741 GCTACCCCAAGCAATATTCAGCAGAGTCAATTTGAAGTTTACGAAGTTGACTGCTT 1800
Db 65337 GCTACCCCAAGCAATATTCAGCAGAGTCAATTTGAAGTTTACGAAGTTGACTGCTT 65396
QY 1801 GTGGCTCAAGCTTTTATTTGTGAAGCCACACCTTTTACAGATGCTGATTAAGCAATTC 1860
Db 65397 GTGGCTCAAGCTTTTATTTGTGAAGCCACACCTTTTACAGATGCTGATTAAGCAATTC 65456
QY 1861 CAGCAACTTTCAGCAGGCTTAAAGTAAATGGCTATGAGATTCCTGAGATTTGTATGATACA 1920
Db 65457 CAGCAACTTTCAGCAGGCTTAAAGTAAATGGCTATGAGATTCCTGAGATTTGTATGATACA 65516
QY 1921 CGGAATTAATTTGGAGATGACTTGGCTGAGAAAGGGAATCTGTGCACTGCTTATAGGC 1980
Db 65517 CGGAATTAATTTGGAGATGACTTGGCTGAGAAAGGGAATCTGTGCACTGCTTATAGGC 65576
QY 1981 AAAGTTGATGAATGCGTATGTGGTTGGCTTGAACAGTGAAGATTCATATATAGGAAT 2040
Db 65577 AAAGTTGATGAATGCGTATGTGGTTGGCTTGAACAGTGAAGATTCATATATAGGAAT 65636
QY 2041 CCAGCTTATTTGTGAGTTTGTGAGAAATTCATATGATGACATGATGATCTCCCT 2100
Db 65637 CCAGCTTATTTGTGAGTTTGTGAGAAATTCATATGATGACATGATGATCTCCCT 65696
QY 2101 GAGCTATGAAATTTGTGAGAACTGTGTGAGAGGTTGTCTTCCAGGTTCAAGAC 2160
Db 65697 GAGCTATGAAATTTGTGAGAACTGTGTGAGAGGTTGTCTTCCAGGTTCAAGAC 65756
QY 2161 ACCAAGATTAATAATTTTAACTCGGGGACTATGATGATGATCTTATGTTGATGATAC 2220
Db 65757 ACCAAGATTAATAATTTTAACTCGGGGACTATGATGATGATCTTATGTTGATGATAC 65816
QY 2221 TTGGAAGAAGTGAAGTGTGAGGTTCTCTTACTGCTGTGCAACTATGAGCAGAG 2280
Db 65817 TTGGAAGAAGTGAAGTGTGAGGTTCTCTTACTGCTGTGCAACTATGAGCAGAG 65876
QY 2281 ATTGAGCCGAGCATGTGAAGGCTAGTATGACGAGCATGCGAAGAGTTTCTCTTCC 2340
Db 65877 ATTGAGCCGAGCATGTGAAGGCTAGTATGACGAGCATGCGAAGAGTTTCTCTTCC 65936

QY 2341 CGCTATACAGATAGAACTTCGCGTGAACCCAGAGATGTGCAAGACAGTGTATGCTGA 2400
Db 65937 CGCTATACAGATAGAACTTCGCGTGAACCCAGAGATGTGCAAGACAGTGTATGCTGA 65996
QY 2401 GATCCTGTGTGAACAATGAGGCGGATGAGGCTGGTGTCTTATTTGAGAAAGCT 2460
Db 65997 GATCCTGTGTGAACAATGAGGCGGATGAGGCTGGTGTCTTATTTGAGAAAGCT 66056
QY 2461 GTAAGACCTCTGAGAACTTTGAGAACTTATGATTAAGCAATTCGAGCTGGGGTCTCAGAG 2520
Db 66057 GTAAGACCTCTGAGAACTTTGAGAACTTATGATTAAGCAATTCGAGCTGGGGTCTCAGAG 66116
QY 2521 AGTAGCCTGATGAACACTATGTTGAAATGTCGGTGTCTGATATGTTTAAAGAGGCAAGT 2580
Db 66117 AGTAGCCTGATGAACACTATGTTGAAATGTCGGTGTCTGATATGTTTAAAGAGGCAAGT 66176
QY 2581 GTGAAGATCTTACTGTGTGTGTGAGCAATTTGAGCTGATTTCACTGTTCAAGCAAGATAT 2640
Db 66177 GTGAAGATCTTACTGTGTGTGTGAGCAATTTGAGCTGATTTCACTGTTCAAGCAAGATAT 66236
QY 2641 TTTCTTAAAGCAGCTCATCTTTTCAACGCAAGATATGTTTTCTTATGAAATCTGAT 2700
Db 66237 TTTCTTAAAGCAGCTCATCTTTTCAACGCAAGATATGTTTTCTTATGAAATCTGAT 66296
QY 2701 GTGCTACCTATGATGATTAATGATGCAATTTTCAATATCTGATGCTCAAAATA 2760
Db 66297 GTGCTACCTATGATGATTAATGATGCAATTTTCAATATCTGATGCTCAAAATA 66356
QY 2761 TGCCTGTTTTGTAGGCTTGAAGACATAGTTCCACTTAATATCATGTGCCAAAGTTGACC 2820
Db 66357 TGCCTGTTTTGTAGGCTTGAAGACATAGTTCCACTTAATATCATGTGCCAAAGTTGACC 66416
QY 2821 AAGATTAACAAGTTGCTGATTAATTTCACTAATATGCTGTGAATTTTTGATCAAA 2880
Db 66417 AAGATTAACAAGTTGCTGATTAATTTCACTAATATGCTGTGAATTTTTGATCAAA 66476
QY 2881 CTGTAGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTTGAATTAAGTGAATAG 2940
Db 66477 CTGTAGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTTGAATTAAGTGAATAG 66536
QY 2941 AGATGCTAGTGTGGCTTTGTCCAACCTTTTCTTCTGATTTTTTCTTTTGCATTT 3000
Db 66537 AGATGCTAGTGTGGCTTTGTCCAACCTTTTCTTCTGATTTTTTCTTTTGCATTT 66596
QY 3001 AGGCTCAGTCAAGCTGACGATTCAGAAAGCATTCGCAAGATGATGCTAGACCTGAGA 3060
Db 66597 AGGCTCAGTCAAGCTGACGATTCAGAAAGCATTCGCAAGATGATGCTAGACCTGAGA 66656
QY 3061 GAATATATGATCCAGTGTGCAAGAAATTAAGTCTGTGGCTTTGGGCTGTATCACCGCAT 3120
Db 66657 GAATATATGATCCAGTGTGCAAGAAATTAAGTCTGTGGCTTTGGGCTGTATCACCGCAT 66716
QY 3121 AGAAATGTTTACAGAGGTGAGGGAATTAATCTCAATCAATCAATTTGTGTGAAGACTGT 3180
Db 66717 AGAAATGTTTACAGAGGTGAGGGAATTAATCTCAATCAATCAATTTGTGTGAAGACTGT 66776
QY 3181 TGACATGATTAATATGCTGTGTGCTTTGTGATTTCTGTTATTAAGTTTGTGATGGC 3240
Db 66777 TGACATGATTAATATGCTGTGTGCTTTGTGATTTCTGTTATTAAGTTTGTGATGGC 66836
QY 3241 GAATGCTGAAGATTTGAGCTGACAGAGCAGCTGAAATCTGGCAGCTTTGGTGTGATG 3300
Db 66837 GAATGCTGAAGATTTGAGCTGACAGAGCAGCTGAAATCTGGCAGCTTTGGTGTGATG 66896
QY 3301 ATTATACATGTTGAACATATCTGTTGACAGTGTGACAGCTGAGAGATGAAACCGGT 3360
Db 66897 ATTATACATGTTGAACATATCTGTTGACAGTGTGACAGCTGAGAGATGAAACCGGT 66956
QY 3361 CTCTGTGGAAGCACTCTGAGAGAGTGTGCTATCTGATTTGTTTCAATCCAGAAA 3420
Db 66957 CTCTGTGGAAGCACTCTGAGAGAGTGTGCTATCTGATTTGTTTCAATCCAGAAA 67016
QY 3421 ACAATGCTACTGATGTCAAGAACTTACACAAAGATTCAGAAAGTTTCTGTGCCAAGTCA 3480

Db	Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Comment
Db	67017	ACAACTGCTACTGATGTGAGAACTTACCAACAGATACGAAAGTTTCTGTGTCATGTCAG	670176							
Qy	3481	GGTGGAAATCACTGAAGGCTCTGTCTTCTGTCATCATATATATCTCATATGTAGATGTC	3540							
Db	67077	GGTGGAAATCACTGAAGGCTCTGTCTTCTGTCATCATATATATCTCATATGTAGATGTC	67136							
Qy	3541	GAGCTTGGGAGATCTCTTTGTTCTGTAAATCTCTCTCTAAATGTAGTGTATTAATGA	3600							
Db	67137	GAGCTTGGGAGATCTCTTTGTTCTGTAAATCTCTCTCTAAATGTAGTGTATTAATGA	67136							
Qy	3601	ACACAAAATATTAAGCTTCTGGGACACCCCTTCTGATCTTAACATTAACATTAAGG	3660							
Db	67197	ACACAAAATATTAAGCTTCTGGGACACCCCTTCTGATCTTAACATTAACATTAAGG	67256							
Qy	3661	GCTACAA 3667								
Db	67257	GCTACAA 67263								
RESULT 4	AY091075	2637 bp	mRNA	linear	PLN 18-SEP-2002					
LOCUS	AY091075									
DEFINITION	Arabidopsis thaliana unknown protein (At5g42480) mRNA, complete cds.									
ACCESSION	AY091075									
VERSION	AY091075.1	GI:20259550								
KEYWORDS	FLI CDNA.									
SOURCE	Arabidopsis thaliana (thale cress)									
ORGANISM	Arabidopsis thaliana									
REFERENCE	1 (bases 1 to 2637)									
AUTHORS	Yamada, K., Banb, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldemich, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Saitou, M., Seki, M., Shim, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.									
TITLE	Arabidopsis Full Length cDNA Clones									
JOURNAL	Unpublished									
REFERENCE	2 (bases 1 to 2637)									
AUTHORS	Yamada, K., Banb, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldemich, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Saitou, M., Seki, M., Shim, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.									
TITLE	Direct Submission									
JOURNAL	Submitted (21-MAR-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA									
COMMENT	RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFI cDNAs (RAFI cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Saitou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.									

The Salk, Stanford, PGSC (SSP) Consortium members carried out the sequencing and annotation of the RAFI cDNAs: Yamada, K., Banb, J., Chan, M.M., Chang, C.H., Chang, E., Dale, J.M., Deng, J.M., Goldemich, A.D., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wu, H.C., Yamamura, Y., Yu, G., Bowser, L., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Kim, C., Lam, B., Lin, J., Meyers, M.C., Miranda, M., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Saitou, M., Seki, M., Shim, P., Southwick, A., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Southwick, A., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGSC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGSC)

contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to Genbank.

Location/Qualifiers

FEATURES

Source

1. .2637

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="5"

/clone="RAF109-76-G11 (R19395)"

/ecotype="Columbia"

/note="This clone is in a modified pBluescript vector (pUC-1) as a BamHI/XhoI insert."

1. .2637

/gene="At5g42480"

1. .113

/gene="At5g42480"

114. .2519

/gene="At5g42480"

/codon_start=1

/evidence=experimental

/product="unknown protein"

/protein_id="AA13895.1"

/db_xref="GI:20259551"

/translation="MRALSHVIGTGLSPFLCRLPATTKLRBSHNTSTTCSAKMADRLSDENFTSDSSSPATRTATTTATLVSPSPSIDRPERHPIDPQVGAQHFLLDGRIRFEARVSKRPQGFSDALISRQLQAACETLSPSRREINELGDBEATVIDRVPDKRGVALCVLQEGEGETIVLRVBLAKELPKSPFKQDVLVNAFLADVSRLDMALDDPDTITGYEFVEBALKLOEGSASLPDRLAIDETLEITPRYVELIGLPLDQDPAARKRLNGLSGVRIIWSVGGASALVGLTRKPMNEALRMTAAEQVDLFLATPSNIPASFEYVEVALALVAQAFICKPHLLDQADKFOQLQAAMEIPLMFLYTRNNWETDFEGERLCAHLIGKVDCEKMLGDSBDQYRNPALVEYLSNRDLNDLDPGLCLBETWLAGVVEPRFDIKDKFKLAGDYDDPMVLSYLERVEVQGSFLAAAAAARIRIAEHVKAAMQLOKVPFSRYTDNRNSAPEKDVQETVESVEADGVNNGVDGEEVGFALAEVRPSENETNDVALIRAGVSSSVDETFEWSVADMLEKASVKILAAQVAVIGLIFSKVYKIKSSSPQRDMVSSSVATIGVSAVDSEALPRMDARTENIIVSWMQKTKSLAFGPDHRIEMLPYLDGRMLKITDPAALTRDQGLAYDPTLKLSTNIVTYSAGSTRALVERATLEESACLSDLVHPENNATDVRIITRYEIVFMSKSMKITEGSVLac"

2520. .2637

/gene="At5g42480"

ORIGIN

3' UTR

Query Match 52.7%; Score 1931; DB 8; Length 2637;

Best Local Similarity 80.4%; Pred. No. 0;

Matches 2616; Conservative 0; Mismatches 5; Indels 632; Gaps 5;

368 GATTAACTTATACTACTCTCAAAATCAAAATTCCTATAACCTTAGACGACCAACAAAGCTC 427

1 GATTAACTTATACTACTCTCAAAATCAAAATTCCTATAACCTTAGACGACCAACAAAGCTC 60

428 TTCAATATGTAAACAGAAACAAAGTTTTTGTAGAGCTTAAAAAGACACTCCCATGGAAG 487

61 TTCAATATGTAAACAGAAACAAAGTTTTTGTAGAGCTTAAAAAGACACTCCCATGGAAG 120

488 CTCTGAGTCACGTGGCATTGGTCTCTCCCATTCCAATTATGCGCATTAACACCGCGCA 547

121 CTCTGAGTCACGTGGCATTGGTCTCTCCCATTCCAATTATGCGCATTAACACCGCGCA 180

548 CGAAGAAAGCTCCGACGTAGCCACAAACCTCTACAACTATATCTGTCGCGACGAAAGGG 607

181 CGAAGAAAGCTCCGACGTAGCCACAAACCTCTACAACTATATCTGTCGCGACGAAAGGG 240

608 CGACCGCTCTCTCTCGGACCTTCAATTTGACACTCCGATTCCTCTCTCTCTCTTGGCCA 667

241 CGACCGCTCTCTCTCGGACCTTCAATTTGACACTCTCCATTCCTCTCTCTCTTGGCCA 300

668 CGGCAACCAACCGGCACTCTGCTCTCTCTGACCACTATTTGATCGTCCGAAAGCC 727

301 CGGCAACCAACCGGCACTCTGCTCTCTCTGACCACTATTTGATCGTCCGAAAGCC 360

728 AGGTCCCAATCCCAATGATTTCTACCAAGGATTTAGAGCTCAAAACATTTCTTAACG 787


```

QY 2948 CTTAGTGTGGCTTTGTCCAACTTTCTTCTGATTTTCTTTGATTTAGGGTCA 3007
Db 2092 -----TAGGGTCA 2099
QY 3008 GTGAGAGCTGACGATTCAGAGCACTTCCAGAAATGATGTAGACTGCAGAGATATA 3067
Db 2100 GTGAGAGCTGACGATTCAGAGCACTTCCAGAAATGATGTAGACTGCAGAGATATA 2159
QY 3068 GATTCAGAGTGGCAGAGATTAAGTCTGTGGCTTTGGGCTTATCACCCGATAGAAATG 3127
Db 2160 GATTCAGAGTGGCAGAGATTAAGTCTGTGGCTTTGGGCTTATCACCCGATAGAAATG 2219
QY 3128 TTTCAGAGGTGAGGAGTAATTAATCAATTCATCAATGTGTGAAATCTGTGGACAT 3187
Db 2220 TTTCAGG----- 2226
QY 3188 GATTATAGTCTGTGCTGCTTTGATTCGTATTATTAAGGTTTGGATGGGCGAATGCT 3247
Db 2227 -----AGCTTTTGATGGGCGAATGCT 2248
QY 3248 GAAAGATTGACCTGACAGAGCACTGAAACTGCGACCTTGGGTTGATTATATAC 3307
Db 2249 GAAAGATTGACCTGACAGAGCACTGAAACTGCGACCTTGGGTTGATTATATAC 2308
QY 3308 ACTGTTGAAACTATCTGTTTGAACAGTGTGACAGTCTCAGAGATGGAACCCGCTCTGCT 3367
Db 2309 ACTGTTGAAACTATCTGTTTGAACAGTGTGACAGTCTCAGAGATGGAACCCGCTCTGCT 2368
QY 3368 GGAAGCAACTCTGAGAGAGTCTGCTCTATCTATCTGATTTGGTTTCATCCAGAAACAATGC 3427
Db 2369 GGAAGCAACTCTGAGAGAGTCTGCTCTATCTGATTTGGTTTCATCCAGAAACAATGC 2428
QY 3428 TACTGATGTGAGAACTTACACACAGATGAGAAAGTTTCTGCTCAGAGTCAAGGTGAAA 3487
Db 2429 TACTGATGTGAGAACTTACACACAGATGAGAAAGTTTCTGCTCAGAGTCAAGGTGAAA 2488
QY 3488 AATCACTGAGAGGCTCTGTTCTGTGATCATATATATCAATATGAGATGCTGAGCTTG 3547
Db 2489 AATCACTGAGAGGCTCTGTTCTGTGATCATATATATCAATATGAGATGCTGAGCTTG 2548
QY 3548 CGAGATTCCTGTTGCTGTAATTCCTCTCTAAGTATGATTTATTAATGAACACAAA 3607
Db 2549 CGAGATTCCTGTTGCTGTAATTCCTCTCTAAGTATGATTTATTAATGAACACAAA 2608
QY 3608 AAAATTAACGTTT 3620
Db 2609 AAAATTAACGTTT 2621

```

RESULT 5
AY221469 2438 bp mRNA linear PLN 05-AUG-2003
LOCUS Arabidopsis thaliana division protein (ARCF) mRNA, complete cds;
DEFINITION nuclear gene for chloroplast product.
ACCESSION AY221469
VERSION AY221469.1 GI:33436353
SOURCE Arabidopsis thaliana (chale crees)
ORGANISM Arabidopsis thaliana
KEYWORDS Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 2438)
Vitha,S., Froehlich,J.E., Koksharova,O., Pyke,K.A., Van Erp,H. and Oesteryoung,K.W.
AUTHORS ARCF is a J-domain Plastid Division Protein and an Evolutionary Descendant of the Cyanobacterial Cell Division protein Ftn2
JOURNAL Plant Cell 15 (8), 1918-1933 (2003)
PUBMED 12897262
REFERENCE 2 (bases 1 to 2438)
Vitha,S., Koksharova,O., van Erp,H., Froehlich,J.E. and Oesteryoung,K.W.

```

TITLE Direct Submission
JOURNAL Submitted (17-JUN-2003) Department of Plant Biology, Michigan State University, 166 Plant Biology Building, East Lansing, MI 48824, USA
FEATURES
SOURCE
  /organism="Arabidopsis thaliana"
  /mol_type="mRNA"
  /db_xref="taxon:3702"
  /ecotype="Wassiljevskij"
  1. 2438
  /gene="ARCF"
  /note="synonym: At5g42480"
  2. 2407
  /gene="ARCF"
  /codon_start=1
  /product="division protein"
  /protein_id="AAQ18646.1"
  /db_xref="GI:33436354"
  /translation="MEALSHVIGIGLSPQCRLEPPATTKLRSHNTSTTICSAKMDRLSDPMTSDSSSPATATTTATTVSLPSIDRPRRHVPIDFYOVIGAOHPLTDGIRRAEARAVSKPQPFSDDALISROILQAACTLSNPSRREYNEGLDDEGATVITDVPMDKVPKALCYLQEGEETIVLRVGBALUKERLPPSKFQDVVLVWALALDYSRDMAALDPDPIITGYEFEBEALKLQEGASLAPDLRAQIDETLEETIPRYVLELLGLPGDDVAAKRLNGLSGVNRLMSVGGASLAVGILTRKFMNEARLPMTPAAQVLDLFAVPSNIPASEFEVEVALALVAOAFIGKKPHILOADPOFOOLOAKWAMEIPLMDNDLRLCTLETLTAGVYPPRPDRKDKFKLGDYDDPPMLSYLERVVGSPDLAAATMARICAEHVKASAMQALQKVPFSRTYDRSAPKQVQETFSVDVGNVNGDGEQGVFAEVRPSENFTNDVAIRAGVSSSVDETTVEMSVLEMLKASVXKILAAVAGISLIFSOKEFLKSSSSFORQDMVSMESDVATIGSVADSEALPRMDARAEIVSKOKIKSLAFQDRIEMLPVLDPRMKITWDRAPAEYRQGLVVDYDTLLKSVSVTVSADOTRALVETLESACLSDLVHPEMNATDVRYTTRVYEVFMSKMKITGSEVLA"
misc_difference 474
  /gene="ARCF"
  /note="results in g1y to glu substitution"
  /replace="a"
misc_difference 1831
  /gene="ARCF"
  /note="results in glu to asp substitution"
  /replace="t"
ORIGIN
Query Match 47.8%; Score 1752.8; DB 8; Length 2438;
Best Local Similarity 79.3%; Pred. No. 0;
Matches 2436; Conservative 0; Mismatches 2; Indels 632; Gaps 5;
QY 480 CATGAAGCTCTGAGTCACTGCGATTTGCTCTCCCAATTCATATATGCGGATTAAC 539
Db 1 CATGAAGCTCTGAGTCACTGCGATTTGCTCTCCCAATTCATATATGCGGATTAAC 60
QY 540 ACCGGCGACGACAAAGCTCGACGTAGCCACACACCTCTACAACTATCTGCTCCGAC 599
Db 540 ACCGGCGACGACAAAGCTCGACGTAGCCACACACCTCTACAACTATCTGCTCCGAC 60
QY 61 ACCGGCGACGACAAAGCTCGACGTAGCCACACACCTCTACAACTATCTGCTCCGAC 120
Db 61 ACCGGCGACGACAAAGCTCGACGTAGCCACACACCTCTACAACTATCTGCTCCGAC 120
QY 600 CAAATGGCGGACCGCTTCTCTCCGATCAATTAACCTCGATCTCTCTCTCTCTCTCT 659
Db 121 CAAATGGCGGACCGCTTCTCTCTCCGATCAATTAACCTCGATCTCTCTCTCTCTCTCT 180
QY 660 CTTGCGACCGCCACACACACCGCACTCTGCTCTCTGCAACATCTATATGATCTGCC 719
Db 181 CTTGCGACCGCCACACACACCGCACTCTGCTCTCTGCAACATCTATATGATCTGCC 240
QY 720 CGAAGCGACGTCCATCCCATTTGATTTCTACAGATATGAGAGCTCAACACATTT 779
Db 241 CGAAGCGACGTCCATCCCATTTGATTTCTACAGATATGAGAGCTCAACACATTT 300
QY 780 CTTAACCGATGGAATCGAAGAGATTCGAAGCTTTCGAACCGCGCAATTCGG 839
Db 301 CTTAACCGATGGAATCGAAGAGATTCGAAGCTTTCGAACCGCGCAATTCGG 360
QY 840 TTTCAGCGACGAGCTTTAATCAAGCCGAGACAGATTTCTCAAGCTCTTGGCAACTCT 899

```


Db 361 TTTCAGCAGCAGCGTTTAATCAGCCGAGACAGATTCCTTCAGCGCTGGCGAACTCT 420
Qy 900 GTCTAATCTCGGCTCTAGAAAGAGTACAATGAGAGCTTCTTGATGATGAAGAAGCTAC 959
Db 421 GTCTAATCTCGGCTCTAGAAAGAGTACAATGAGAGCTTCTTGATGATGAAGAAGCTAC 480
Qy 960 AGTCATCACTGATGTTCTTGGGATTAAGTAATTCGATTCGGAATTAATAAGTTTCTT 1019
Db 481 AGTCATCACTGATGTTCTTGGGAT----- 505
Qy 1020 CGTTTAATTCATGAAATGGAATTAAGAAAGAACTTTTATCTAGTGAAGTTCTCGGG 1079
Db 506 -----AAGTTCTCGGG 518
Qy 1080 CTCTCTGTGATTTSCAAGAAGGTGTGAGACTGAGATAGTTCTTCGGGTTGTGAGGCTC 1139
Db 519 CTCTCTGTGATTTSCAAGAAGGTGTGAGACTGAGATAGTTCTTCGGGTTGTGAGGCTC 578
Qy 1140 TGTCTAAGAGAGGTTGCTTAAGTCGTTTAAAGCAAGATGTGTTTAAATTAAGCGCTTG 1199
Db 579 TGTCTAAGAGAGGTTGCTTAAGTCGTTTAAAGCAAGATGTGTTTAAATTAAGCGCTTG 638
Qy 1200 CGTTTCTCGATGTTCTGAGAGGATGCTATGAGATTCACCTCGATTTTATTACTGTT 1259
Db 639 CGTTTCTCGATGTTCTGAGAGGATGCTATGAGATTCACCTCGATTTTATTACTGTT 698
Qy 1260 ATGAGTTTGTGAGAGACTTTGAAGCTTTTACAGATGATTTGACTTGTGTTAAATTT 1319
Db 699 ATGAGTTTGTGAGAGACTTTGAAGCTTTTAC----- 731
Qy 1320 GACGAGCGTTGGCTTTAATAAGAACTTTCGATTTGATTAATGAGTCTGTGT 1379
Db 732 ----- 731
Qy 1380 AGGAGGAAGAGCAATGAGCTTGCACCGATTTAAGTGCAAAATGATGAGACTTTGG 1439
Db 732 AGGAGGAAGAGCAATGAGCTTGCACCGATTTAAGTGCAAAATGATGAGACTTTGG 791
Qy 1440 AAGAGATCACTCCGCTGATGTTCTTGAAGCTACTTGGCTTACCGCTTGTGATGATTAAG 1499
Db 792 AAGAGATCACTCCGCTGATGTTCTTGAAGCTACTTGGCTTACCGCTTGTGATGATTAAG 851
Qy 1500 CTGCGAAAGAATTAATGTTTAAAGCGGTGTGCGAAATATTTGTGTTCTGTTGAGAGAG 1559
Db 852 CTGCGAAAGAATTAATGTTTAAAGCGGTGTGCGAAATATTTGTGTTCTGTTGAGAGAG 911
Qy 1560 GTGAGAGATCAAGCTTGTGTTGGGGTTTGAACCCGTGAGAACTTATGAAGAAGCGCTTT 1619
Db 912 GTGAGAGATCAAGCTTGTGTTGGGGTTTGAACCCGTGAGAACTTATGAAGAAGCGCTTT 971
Qy 1620 TACGAATGACAGCTGTGAGAGGATATACAGTTTATGATTAACCTTTTAAATTTCTTAGC 1679
Db 972 TACGAATGACAGCTGTGAGAG----- 992
Qy 1680 ATGATATTAACCTTAGGTTTCTCATTTTAATGATGTTGTGTGAGGTTGATCTTTTGT 1739
Db 993 -----AGGTTGATCTTTTGT 1008
Qy 1740 AGCTACCCCAAGCAATTTTCAAGAGAGTCAATTTGAAGTTTACGAATTTGCACTTGTCT 1799
Db 1009 AGCTACCCCAAGCAATTTTCAAGAGAGTCAATTTGAAGTTTACGAATTTGCACTTGTCT 1068
Qy 1800 TGTGCTCAAGCTTTTATGTTGAAGCAACCTTTTACAGGATCTGATGAACAAT 1859
Db 1069 TGTGCTCAAGCTTTTATGTTGAAGCAACCTTTTACAGGATCTGATGAACAAT 1128
Qy 1860 CCAAGCAATTTCAAGAGCTTAAGGTAATGCTATGAGATTCCTGCAATGTTGATGATAC 1919
Db 1129 CCAAGCAATTTCAAGAGCTTAAGGTAATGCTATGAGATTCCTGCAATGTTGATGATAC 1188
Qy 1920 ACGGAATTAATTTGGAGATAGACTTCGGTCTTAAGAAAGGGGACTGTGCACTGCTTATAGG 1979
Db 1189 ACGGAATTAATTTGGAGATAGACTTCGGTCTTAAGAAAGGGGACTGTGCACTGCTTATAGG 1248

Qy 1980 CAAAGTTGATGAATGCCGTATGTGTTGGCTTGAACAGTGAAGATTCACATATAGGAA 2039
Db 1249 CAAAGTTGATGAATGCCGTATGTGTTGGCTTGAACAGTGAAGATTCACATATAGGAA 1308
Qy 2040 TCCAGCTATTTGAGAGTTGTTTGTGAGAAATTCAAATCGTATGACATATGATCTCCC 2099
Db 1309 TCCAGCTATTTGAGAGTTGTTTGTGAGAAATTCAAATCGTATGACATATGATCTCCC 1368
Qy 2100 TGAATATGCAAAATTTGTGAGAACTGTGTGGAGGGGTTGTCTTCTCTAGTTCAGAGA 2159
Db 1369 TGAATATGCAAAATTTGTGAGAACTGTGTGGAGGGGTTGTCTTCTCTAGTTCAGAGA 1428
Qy 2160 CACCAAAAGATAAAAATTTAACTCGGGGACTATATGATATCTATGTTTGAATTA 2219
Db 1429 CACCAAAAGATAAAAATTTAACTCGGGGACTATATGATATCTATGTTTGAATTA 1488
Qy 2220 CTTGAAAAGTGAAGTGAAGTTCAAGGTTCTCTTTAGCTGTGCTGCAACTATGCGCAG 2279
Db 1489 CTTGAAAAGTGAAGTGAAGTTCAAGGTTCTCTTTAGCTGTGCTGCAACTATGCGCAG 1548
Qy 2280 GATTGAGCCGAGCATGTGAAAGCTATGTCTATGACAGGCACTGCAAAAGTTTCTTC 2339
Db 1549 GATTGAGCCGAGCATGTGAAAGCTATGTCTATGACAGGCACTGCAAAAGTTTCTTC 1608
Qy 2340 CCGCTATACAGATAGAAACTCGGCTGAAACCAAGATGTGCAAGAGACAGTGTTAAGTGT 2399
Db 1609 CCGCTATACAGATAGAAACTCGGCTGAAACCAAGATGTGCAAGAGACAGTGTTAAGTGT 1668
Qy 2400 AGATCTGTGTTGAACAATGTAGCCGCTGATGTGAGACCTGTGTCTTTATTGCAAGAC 2459
Db 1669 AGATCTGTGTTGAACAATGTAGCCGCTGATGTGAGACCTGTGTCTTTATTGCAAGAC 1728
Qy 2460 TGTAAAGCCCTCTGAAAACCTTTGAACATAATGATTAATGCAATTCGAGCTGGGCTCAGA 2519
Db 1729 TGTAAAGCCCTCTGAAAACCTTTGAACATAATGATTAATGCAATTCGAGCTGGGCTCAGA 1788
Qy 2520 GAGTAGGTTGATGAACATACTGTGAATGTCCGTTGCTGATATGTTAAAGAGGCAAG 2579
Db 1789 GAGTAGGTTGATGAACATACTGTGAATGTCCGTTGCTGATATGTTAAAGAGGCAAG 1848
Qy 2580 TGTGAAGATCTGAGCTGTGTGTGTGCAATTTGACATGATTTTCACTGTTACGCGAAGTA 2639
Db 1849 TGTGAAGATCTGAGCTGTGTGTGTGCAATTTGACATGATTTTCACTGTTACGCGAAGTA 1908
Qy 2640 TTTTCTTAAAGAGAGCTCATCTTTTCAAGCAAGATATGTTCTTCTATGGAATCTGA 2699
Db 1909 TTTTCTTAAAGAGAGCTCATCTTTTCAAGCAAGATATGTTCTTCTATGGAATCTGA 1968
Qy 2700 TGTGCTACCATGATGATATTAATGATGCAATTTTCAATATATCTGCAATTTGCTCAAAAT 2759
Db 1969 TGTGCTACCA----- 1979
Qy 2760 ATGCTGTTTGTGAGCTAAGAACATATGTTCCACTTAATATATGTCCTCAAAAGTTGTAC 2819
Db 1980 ----- 1979
Qy 2820 CAAGATTAACAAGTTGTGATGAATTTCACTAATTAATGCTGTGATTAATTTTGTATCA 2879
Db 1980 ----- 1979
Qy 2880 ACTGTAGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTTGAATTAACGTAGATTA 2939
Db 1980 ----- 1979
Qy 2940 GAGATTGCTTATGTGAGCTTTGTCCAATTTTCTTCTTGAATTTTCTTTTGTGATTT 2999
Db 1980 ----- 1979
Qy 3000 TAGGGTCAATCAGAGCTGACGATTTCAAGAGCACTTCCAGAAATGATGCTAGGACTGAC 3059
Db 1980 TAGGGTCAATCAGAGCTGACGATTTCAAGAGCACTTCCAGAAATGATGCTAGGACTGAC 2039

OY		3660	AGATATTGATTCGAAGTGGCGAAGAAGTTAGTGCTGTGGSCITTTGGGCCTGATCAACGCCA	3119
Db		2040	AGAAATAAGTAATCCAAAGTGGCGAAGAGATTAAGTCTGTGGCITTGGGGCTGATCACGCCA	2099
OY		3120	TAGAAAATGTTAACGAGAGGTGAGGGAAATTAAATCTACAATTCAATCAAATTGTGTGAAAACTG	3179
Db		2100	TAGAAAATGTTAACG-----	2114
OY		3180	TTGGACATGATTAATGACTGTGGTGCCCTTGTTGAATCTGTATTATTAAGTTTTGGAATGGG	3239
Db		2115	-----AGGTTTTGGAATGGG	2128
OY		3240	GGAATGCTGGAAGATTTGGCATCTGACACGACAGCTGAAACTGGCGAGCTTGGSTGGTTAT	3299
Db		2129	GGAATGCTGGAAGATTTGGCATCTGACACGACAGCTGAAACTGGCGAGCTTGGSTGGTTAT	2188
OY		3300	GATTATACACTGTTGAAACTATCTCTGTGACAGTGTGACAGTCTGACGAGATGAAACCCT	3359
Db		2189	GATTATACACTGTTGAAACTATCTGTGACAGTGTGACAGTCTGACGAGATGAAACCCT	2248
OY		3360	GCTGTGTTGGAAGCACTCTGAGAGAGTCTGTGTTGTTATCTGATTTGATTGATTCAGAA	3419
Db		2249	GCTGTGTTGGAAGCACTCTGAGAGAGTCTGTGTTGTTATCTGATTTGATTGATTCAGAA	2308
OY		3420	AACAAATGCTACTGATGTCAGAACTTACACAAACAAGATPACGAAGTTTCTGTGCCAACTCA	3479
Db		2309	AACAAATGCTACTGATGTCAGAACTTACACAAACAAGATPACGAAGTTTCTGTGCCAACTCA	2368
OY		3480	GGGTGGAANAATCATCTGAAGGCTGTGTTCTTGCATCATPAATPATCTCATATGATGATGTC	3539
Db		2269	GGGTGGAANAATCATCTGAAGGCTGTGTTCTTGCATCATPAATPATCTCATATGATGATGTC	2428
OY		3540	TGAGCTTGGCG	3549
Db		2429	TGAGCTTGGCG	2438
RESULT_6 AY150490	DEFINITION	Arabidopsis thaliana unknown protein (At5g42480) mRNA, complete cds.		
Accession	Version	KeyWords	SOURCE	ORGANISM
AY150490	AY150490.1	GI:23297711	FPLI CDNA.	Arabidopsis thaliana (chale cress)
REFERENCE	AUTHORS			
				Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosoids; eurosids II; Brassicales; Brassicaceae; Arabidopsia. 1 (bases 1 to 2436) Yanada,H.K., Chan,M.M., Chang,C.H., Dale,J.M., Huan,V.W., Lee,J.M., Quach,H.L., Tang,C.C., Toriumi,M., Wallender,E.R., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Inohide,I., Jones,T., Kamiya,A., Kawai,I., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimizu,P., Southwick,A.A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W., Ecker,J.R. and Theologis A. Arabidopsis Open Reading Frame (ORF) Clones Unpublished 2 (bases 1 to 2436) Yanada,H.K., Chan,M.M., Chang,C.H., Dale,J.M., Huan,V.W., Lee,J.M., Ouach,H.L., Tang,C.C., Toriumi,M., Wallender,E.R., Wong,C., Mu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Inohde,I., Jones,T., Kamiya,A., Kawai,I., Kim C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki M., Shimu,P., Southwick,A.A., Tripp,M.G., Wu,T., Shinzaki,K., Davis,R.W., Ecker,J.R. and Theologis A. Direct Submission Submitted (13-SEP-2002) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN

The Salk, Stanford, PGSC (SSP) Consortium members constructed and sequenced the PUN1 (ORF) clones using the RAPL cDNA: Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Huan, V.W., Lee, J.M., Quach, H.L., Tang, C.C., Tortum, M., Wallender, E.K., Mong, C., Wu, H.C., Yu, G., Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M., Palm, C.J., Shim, P., Southwick, A., Tlup, M.G., Wu, T., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGSC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGSC) contributed equally to this work as pIs.

Annotation is based on the January 2002 version of the Arabidopsis genome submitted to GenBank.

location/Qualifiers

1. .2436

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="5"

/clone="U19395"

/ecotype="Columbia"

/note="This clone is in PUN1 51."

1. .2436

/gene="At5g42480"

1. .2406

/gene="At5g42480"

/codon_start=1

/evidence=experimental

/product="unknown protein"

/protein_id="AA12907.1"

/db_xref="GI:23297712"

/translation="MEALSHVGTGLSPFQCLRLPATYTKLRSHNSTQSCAKMADRLSDSFIDSDSSSFATRTATTTATLVSPPSIDPREAHYPIPIDFQVIGAQCHPLITDGLRRARFARVSKPQGFSDALISRQLQAACEITLSPRSRRRENEGLDDEKNTVIDPMDKVPKALCVLQEGEETIVLVEBALKEKLPFSKQDVVLAVALDPAFLDNRMAALDDPPIGTGYEVEBALKLOEGASLAPDLQAIIDETLEITPRVYELIGLPGIDVAAKRLNGLSGVRIIIMSVGGASVAVGLTRKPMNBAFLRMTAAEVDMLFATPSNIPASREYVEVALVAQAATGKPHLDDADQFOOLQOAKYAMMIIPAMLYIDRNWBIIDFGLEKGLCALIGKNDCEKMIJLDSBDSQYRPAIYERVENSRDADALPGLIKGLHETWAGVPPRPDIDKDKFLGDIYDDPMVLSYERVENVQSGPLAAAAAMRIAGEHYKASAMQALQKPPSRITDRNSAEKQDQVTFSEVDGNNVGDGGEVFAIABVRPSENDETIDYAIIRAGVSSVDETVEVSADMLKEASVYKILAAGVATGILTSPOKXPLKSSSFQRDMVSMESDVATTGSRVADSSEALPRMDARPAENIVSKRQKTKSLAFGEEDRIEMLPYVDGRMLKITDRPAEFGALGVYDTLLKLSVSVTSYADSTRALVEATLEESACLSDLVHPENNATIDVRYTTRVEVMSKSMKITEGSLAS"

2407. .2436

/gene="At5g42480"

2433. .2434

/gene="At5g42480"

/note="compared to genomic sequence"

/replace="ctg"

ORIGIN

Query Match 47.6%; Score 1744.6; DB 8; Length 2436;

Best Local Similarity 79.2%; Pred. No. 0;

Matches 2429; Conservative 0; Mismatches 4; Indels 632; Gaps 5;

481 ATGGAAGCTCTGAATCAGTCGCGAATGGATGCTCTCCCATTCGAATTATGCGGATTACCA 540

1 ATGGAAGCTCTGAATCAGTCGCGAATGGATGCTCTCCCATTCGAATTATGCGGATTACCA 60

541 CCGCGCAGCAGCAAAAGCTCCGACGTAGCCACAACAACCTCTACAACTATCTGCTCGGCAGC 600

Db 61 CCGCGCAGCAGCAAAAGCTCCGACGTAGCCACAACAACCTCTACAACTATCTGCTCGGCAGC 120

601 AAATGGCCGACCGCTCTTCTCTCCGACTTCAATTTCACCTCCATTCTCTCTCTCTCTCC 660

QY	2821	AGAGTAAACAAGTTGCTGAGTAATTTCACTAATTAATGCTGTAATTTTGTGATCMAA	2880
Db	1979	-----	1978
QY	2881	CTGTAGACAGAAATGTAATTTCACTCTCAACATTTCTGTTAGAAATACAGAGATTAG	2940
Db	1979	-----	1978
QY	2941	AGATGCTTAACTGCTGCTTGTCCAACTTTCTTCTGATTTTCTTTTGTGATTT	3000
Db	1979	-----	1979
QY	3001	AGGCTCAGTCAGACGTCGACATTCAGAGACACTTCCAGAAATGATGCTAGACGTCAG	3060
Db	1980	AGGCTCAGTCAGACGTCGACATTCAGAGACACTTCCAGAAATGATGCTAGACGTCAG	2039
QY	3061	GAATATAGTATCCAGGTCGAGAGATTAAGTCTCTGCTTTGGGCTGATCACCAGAT	3120
Db	2040	GAATATAGTATCCAGGTCGAGAGATTAAGTCTCTGCTTTGGGCTGATCACCAGAT	2099
QY	3121	AGAAATGTTACCAAGGTCGAGAGATTAATCTACATTCATCAATTTGTGAAAACCTGT	3180
Db	2100	AGAAATGTTACCAAGGTCGAGAGATTAATCTACATTCATCAATTTGTGAAAACCTGT	2113
QY	3181	TGACATGATTAATAGTCTGCTGCTGCTTGTGATTTCTGTTATTAAGTCTTTGATGGGC	3240
Db	2114	-----AGTCTTTGATGGGC	2128
QY	3241	GAATGCTGAAGATTGGAAGTCAGAGACGTCGAAATGCGCAGCTTGGTGGTTATG	3300
Db	2129	GAATGCTGAAGATTGGAAGTCAGAGACGTCGAAATGCGCAGCTTGGTGGTTATG	2188
QY	3301	ATTATACACTGTTGAAACTATCTGTTGACAGTGTGACGCTTCAGCAGATGAAACCCGTG	3360
Db	2189	ATTATACACTGTTGAAACTATCTGTTGACAGTGTGACGCTTCAGCAGATGAAACCCGTG	2248
QY	3361	CTCCTGTTGAAGAACCTCTGAGAGAGTCTGCTGCTATCTGATTTGGTTCATCAGAAA	3420
Db	2249	CTCCTGTTGAAGAACCTCTGAGAGAGTCTGCTGCTATCTGATTTGGTTCATCAGAAA	2308
QY	3421	ACATGCTACTGATGTGAGAACTTACACACACACACACACACACACACACACACACAC	3480
Db	2309	ACATGCTACTGATGTGAGAACTTACACACACACACACACACACACACACACACACAC	2368
QY	3481	GGTGAAGAAATACCTGAGAGGCTCTGTTCTGATCATTAATATCTATATGACATGCT	3540
Db	2369	GGTGAAGAAATACCTGAGAGGCTCTGTTCTGATCATTAATATCTATATGACATGCT	2428
QY	3541	GAGCT 3545	
Db	2429	GAGCT 2433	

RESULT 7
AP004885/c
LOCUS AP004885 150462 bp DNA linear PLN 15-SEP-2004
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2,
PAC clone: P0575F10.
ACCESSION AP004885
VERSION AP004885.3 GI:41053009
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharitoidae; Oryzaceae; Oryza.
REFERENCE 1
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 2, PAC
clone: P0575F10
JOURNAL Published Only in Database (2002)
REFERENCE 2 (bases 1 to 150462)
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.

TITLE JOURNAL COMMENT

Direct Submission
Submitted (20-MAR-2002) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: sasaki@nias.affrc.go.jp, URL: http://exp.dna.affrc.go.jp/
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
On Jan 21, 2004 this sequence version replaced gi:38142429.
Genes were predicted from the integrated results of the following:
GENSCAN (http://CCR-081.mt.edu/GENSCAN.html), FGENESH
(http://www.softberry.com/), Genemark (http://opal.biology.gatech.edu/Genemark/), Glimmer
(http://www.tigr.org/cdb/glimmer/glimmer.html), RiceHMM
(http://exp.dna.affrc.go.jp/RiceHMM/), SplicePredictor
(http://bioinformatics.laastate.edu/cgi-bin/sp.cgi), sim4
(http://globin.cse.psu.edu/html/doc/sim4.html), gap2
(http://www.tigr.org/software/glimmer/), BLASTN and BLASTX. The
genomic sequence was searched against NCBI NonRedundant Protein
database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA
sequence database at RGP or DBJ. Protein homologues of the coding
regions were searched against NCBI NonRedundant Protein database
with BLASTP. ESTs represent the identified cDNA sequences using
BLASTN with the corresponding DBJ accession no. and RGP clone ID.
Full-length cDNAs represent the identified cDNA sequences using
BLASTN with the corresponding DBJ accession no.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with full-length cDNA or
EST homology (covering almost the entire length of partial
sequence) is classified as an 'unknown' protein. A gene predicted
by two or more gene prediction programs is classified as a
'hypothetical' protein according to IRSP standard. A gene
predicted by a single gene prediction program is also classified as
a probable 'hypothetical' protein and is included as a
miscellaneous feature of the sequence.
The orientation of the sequence is from T7 to SP6 of the PAC clone.
This sequence of P0575F10 clone has an overlap with OJ1020.C02
(DBJ: AP004078) clone at 5' end and with P0482F12 (DBJ: AP005311)
clone at 3' end. Detailed information on overlap and assembly
quality together with annotation of this entry is available at
http://exp.dna.affrc.go.jp/GenomeSeq.html.
Location/Qualifiers

FEATURES
source
1. 150462
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:3947"
/chromosome="2"
/clone="P0575F10"
complement(4987..8112)
/gene="P0575F10.1"
complement(5003..5412,5520..5679,5764..5914,
6096..6247,6738..6859,7864..8112)
/note="supported by full-length cDNA(s): AK106553"
complement(5003..5412,5520..5679,5764..5914,
6096..6247,6738..6859,7864..8112)
/gene="P0575F10.1"
/note="supported by full-length cDNA(s): AK073059"
complement(5003..5412,5520..5679,5764..5914,
6096..6247,6738..6859,7864..8112)
/note="contains EST(s): AU070150(B61534)
contains full-length cDNA(s): AK106553, AK073059"
/codon_start=1
/product="putative eukaryotic translation initiation
factor 3, subunit 1 alpha, 35kDa"
/protein_id="P0575F10.1"
/db_xref="GI:41053010"
/translation="MEDWDSDDPQPAVPSAKAEPLKSKMADEVEDDYKESWEEREE
EKRPPEVETKAPKPSGKAGKQQAQSSSEVQDEALDDPBLEKRLQRLVEAD
FKSTTELFKKKDGSEKSLDTFIPKSESDFAVEYALINIKKRPYKPSHYMGLKNVNR
LSMNLKQADAKDQISSIAAIANERIKAEKAAAGKKKQAKKQLHIENKDDDFIPG

gene KGNFDDPDEYDEM"
/complement (10186, .13754)
/gene="P0575F10.2"
/complement (join(<10186, .10479,10759, .10892,11127, .11991,
12089, .12349,12515, .12712,12843, .13073,13455, .>13754))
/gene="P0575F10.2"
/note="start and end point are not identified"
/complement (join(10186, .10479,10759, .10892,11127, .11991,
12089, .12349,12515, .12712,12843, .13073,13455, .13754))
/gene="P0575F10.2"
/note="contains EST(s) : AU058443 (E51136), AU183658 (E51136)"
/codon_start=1
/product="placitid division protein precursor"
/protein_id="BAD07942.1"
/db_xref="GI:41053011"
/translation="MEGFHNILARPNAGAPAFSLPRPRPRRRPPHPSACRAAS
WABRLPADPHILPTTAPSDPSAPAPAPASAPVPLPDPARERSLPIQVPPYKTL
GAEHFLDGTIRRAERAKRPPQYGTDLVRRQMQLADHTLNQKSRQYDA
LSEHREALTMIDAMDREAGALAVLTGELLDLRPKPKQDVILAMALAYDLR
DAMAASPPDIYGCCEVERALKLQEDGASNLADLSQIDLEETTPCVLESL
PIDEHKKOEGLQAGANILMSVGRGSIATVGGFSREAFLMNEAFPMRISIEOMDF
KPTNSIIPPEWFEIYNVALVAQAIIISKRPQITMADLFEOLQKNIGSHAYDNE
MDALERAPCSLIVGVSKRMWIGINESPPVADPLRIETVNSSISENDILPGL
CKLETWLIIEVPRSRDTRGMQRLDIYDDEVSLYLERMEGGASHALAAAIK
LGQATVATGIVSNAIQAFKVFPLLEQDRSMENTKQSGSYLENFQENAPAD
SRNALKTIISAGALFALAVIAGAKYLRKRLSIRSHGSVAANSVSDTDPALDE
DPVHPRMDAKLADIYRKQSIKSKALGEPHSVASIQEVLQGNMLKVTWDRAEIER
HGMWEYTLSDVTISITISIDGRRAVTEATIDBAQGLTVTEPRNNDSDTKYTRY
EMAFKLGKWKITGEGAVLKS"
/complement (16401, .17341)
/gene="P0575F10.3"
/complement (join(<16401, .16739,16841, .>17341))
/note="start and end point are not identified"
/complement (join(16401, .16739,16841, .17341))
/gene="P0575F10.3"
/note="similar to Arabidopsis thaliana chromosome 3,
At3g61870"
/codon_start=1
/product="hypothetical protein"
/protein_id="BAD07943.1"
/db_xref="GI:41053012"
/translation="WMAPIAAVAHAIITCSASKNPSPSARQQOQTTTATNGSPA
LPSILRTTAAATAALALAPDPAALAGGEGILGGSVALHLPLWGGIFAYTLWG
YLGWOMRVRTIODEINELKKQLPAAALAPPAVAGDSSSSSSPSPKSPVVI
KIDELTERKKLIGSPRDRHFNAGSILLGIVTESVGLALNTMFRGKLPQPHFA
GAATITVMAAALVPRAMOKGNERYASLHIALNAIVLPLTQIPLGLIVGVFERT
TWP"
/complement (18499, .18942)
/gene="P0575F10.4"
/complement (<18499, .>18942)
/gene="P0575F10.4"
/note="start and end point are not identified"
/complement (18499, .18942)
/gene="P0575F10.4"
/note="contains EST(s) : AU184851 (R10374), AU184850 (R10374)"
/codon_start=1
/product="putative BF-hand Ca2+-binding protein CCD1"
/protein_id="BAD07944.1"
/db_xref="GI:41053013"
/translation="MNHQOIRSTTAAEQEASAGGGGSEYEDLMFVMAGRIGAEGL
LSEIRAGFLIADPARGAITAEISLRSAAYVLG/GGGGGEVTEEAANAVREGDQD
GAISEAFVCLMYLSPGIMDAGMLEAIVDELRLSLPPPPA"
/gene="P0575F10.5"
/join(<21937, .22225,22326, .22648,22736, .23049,
23393, .>23552)
/gene="P0575F10.5"
/note="start and end point are not identified"
/join(21937, .22225,22326, .22648,22736, .23049,23393, .23552)
/gene="P0575F10.5"
/note="similar to Arabidopsis thaliana chromosome 5,
At5g24330"

/codon_start=1
/product="hypothetical protein"
/protein_id="BAD07945.1"
/db_xref="GI:41053014"
/translation="MGPATPLRRRTTAPATRAEGSGGDDGDDVRCAGCSGEA
AEILLDCDRGHICRLRPLRPVAPAGMFCSCSPSHSKSHAKKQKLPSPDDPAR
TKIVDFKIRGPAALAAAESEBGRKRRKRGIRLVSKKRLPLPSPDDPAR
LRQMASLATLVTGAVFNSNELTYVGMAPRANRALESGAQVLPKEDVETLNCK
RMAAGPSPPLVVDVBEVTEARFIDLTITIEYVDVLTTRRDDSDSNMT
LISAATPSPRLVCPDKRSNIARFINGINHTDGRKKQKLCVRFVVGCECHLVLA
NRDISGERLYDYNGSEHEYPTHPV"
/complement (join(24479, .24817,24913, .25266,25398, .25633,
25720, .25832,27054, .27185,27995, .28148))
/gene="P0575F10.6-1"
/complement (join(24479, .24817,24913, .25266,25398, .25633,
25720, .25832,27054, .27185,27995, .28148))
/gene="P0575F10.6-1"
/note="supported by full-length cDNA(s) : AK121372"
/complement (join(24491, .24817,24913, .25266,25398, .25633,
25720, .25832,27054, .27185,27995, .28147))
/gene="P0575F10.6-1"
/note="supported by full-length cDNA(s) : AK062255"
/complement (24563, .28147)
/gene="P0575F10.6-2"
/complement (24563, .28147)
/gene="P0575F10.6-2"
/note="contains full-length cDNA(s) : AK065148
/note="contains full-length cDNA(s) : AK065148
non-coding transcript
probably inactive due to including stop codon(s) in CDS"
/complement (join(24787, .24817,24913, .25266,25398, .25633,
25720, .25832,27054, .27159))
/gene="P0575F10.6-1"
/note="contains EST(s) : AU031017 (E60571), AU093466 (E60571)
contains full-length cDNA(s) : AK121372, AK062255"
/codon_start=1
/product="putative arginine/serine-rich splicing factor"

Query Match 8.7%; Score 317.6; DB 8; Length 150462;
Best Local Similarity 49.3%; Pred. No. 3.3e-64;
Matches 1459; Conservative 0; Mismatches 1299; Indels 202; Gaps 16;

Qy 744 TGATTTCTACAGATATTAGAGACTCAACACATTTCTTAACCGATCGAATCGAAGAGC 803
Db 13086 TGGCTTGTCTCAGAGTTCTAGGGCAGAGCCACATTTCTTGGGATGCTCAGAGAGGC 13027
Qy 804 ATTGGAAGCTAGGGTTTCGAACCGCGCAATTCGGTTTCAGGACGACGCTTTAATCAG 863
Db 13026 GTTCGAGGACGGAATGACCAAGCCAGCTATGGCTACAGACGGAATGCTTTGTGG 12967
Qy 864 CCGAAGACAGATTCCTCAAGCTCTTCGAACTGTGTCTAATCTTCGGTCTAGAAAGA 923
Db 12966 TCGTCGACAAATGCTGAGATTGCCATGACATCTCATGAAACGAACTCCGCACTCA 12907
Qy 924 GTTCAATGAAGGTTCTTGTATGATGAAGAGCTACAGTATCACTGATGTTCTTGGGA 983
Db 12906 GTATGATCGGCCCTTCTGAGAACCGTGAAGAGCTCTCACATGGATATTCCTGGGA 12847
Qy 984 TAAGTAATTTGCAATTCGGAATATTAAGTTTCT-----TCGT 1022
Db 12846 CAAGGTGAGGCTTTAGGGGACCAAGAGGTGTGTTGATCTTGCAATATAGACTAAT 12787
Qy 1023 TTTAATTTCAATGATTTGATTAAGAAAGAACTTTATCTAGTAAAGTTCTTGGGCTC 1082
Db 12786 GTATTAATCTCTCTGCTGTTAATATATATGAGGTGTCTTGGAAGCAGGTTCCAGGTGT 12727
Qy 1083 TCTGTGATTTGCAAGAGGTGTGAGACTGAGATAGTTCTTCCGGTTGGTGGAGCTCTGC 1142
Db 12726 TGTGTGGCTTTCAGAGAGCTGCGGAGGACCTTCTGCTGTGTAATCTGAGAAACAGTTGC 12667
Qy 1143 TTAAGGAGAGGTTGCTTAAGTGTGTTTAAGCAAGATGAGTTTATAGTATGCGCTTGGCT 1202
Db 12666 TCTGATCGGCCACCAAGCGCTTCAAGGACGATGAGTGTGCTAGGATGCTTGGCTT 12607

QY	1203	TTCTCGATGTCGAGGGATGCTATGCGCATTTGGATTCACCTGATTTTATTTACGCGTTAAG	1262
DB	12606	ATGAGCATCTATCAAGGGATGCTTATGGCAGCAGACCCTCCAGATGTAAATTGGCGCTGGCG	12547
QY	1263	AGTTTGTTAGAAAGCTTTGAAGCTTTTACAGG-----TAG	1298
DB	12546	AGGTGCTCGAAGGGCTCTCAAGCTCTTGCGAGGTGAACGTGTTGCTTTTACTCCCTAC	12487
QY	1299	TTTGACTTGCCTTGGTATTTTGAAGACGTTGGCTTTATTAAGAACTTTCCTGATTTG---	1355
DB	12486	TTTTTTCATGCGCTTATATTTTGAATTAAGATTAAGATTAAATTAACCTGTGAAAGATGCTTT	12427
QY	1356	-----ATACTTGTATTGAGCTTGG-----	1376
DB	12426	GGGACACATGGGCTATATTTTATTTAAATGAAGTGGTATTTCTTGAAAGCAATATTATATA	12367
QY	1377	-----TGTAGAGGAGAGAGACAGTATGCTTGACCGGATTTATCGGACAA	1424
DB	12366	GGAACTTCATACCTGTAGAGAGATGAGAGACAGATCTGGACCTGATCTGCTTTCCACAG	12307
QY	1425	TTGATGAGACTTTTGAAGAGATCACTCCGCGTTATGTCTTGAGCTACTTGGCTTACCGC	1484
DB	12306	TTGATGAAGCTCTGAGAGAGATTACACTCGTGCTGTATTTGAGAGCTTCTCCCTTCCAA	12247
QY	1485	TTGGTGATGATTACCGCTGCGAAAAGACTAAATGCTTTAAGCGGTGGCGGAATATTTGT	1544
DB	12246	TTGACACAGACATCATTAAGAAAGCGCCAAAGGGCTTCAAGGTGGGAGAAACATTTTGT	12187
QY	1545	GGTCTGTTGAGAGAGGTGGAGCATCACTCTGTGTGGGGGTTTGACCCGGAGAACTTAA	1604
DB	12186	GGAGCGTTGGCAGAGAGAGATTGCTTACCGTTGAGAGAGATTTCTCGGAAGCCTTCA	12127
QY	1605	TGAATGAGGCGCTTTTAAACGAATGACAGCTGCGAGCAGT-----ATACAGTTTAA	1655
DB	12126	TGAACGAGGCTTTTGAAGATGACATCAATGAACAGGTTTCGACATGTTGTTTTAT	12067
QY	1656	ATACCTTTTAAATTTCTTATAGCATATATACTTTAGGTTTTCTCATTTAAATGATGT	1715
DB	12066	GAAACAGCTGCGCTTATATTTATGACAGAGACATCTTTGAAAACAATGTAACTCCCTTTC	12007
QY	1716	TGT-----GTGTATGTTGATCTTTTGTATGATACCCAGCAATATTCAGAGAGTCA	1770
DB	12006	TGTTAACACTCCAGATGGAATTTCTTTTCAAAAACCGAATGACATCTCTCTGATATG	11947
QY	1771	TTTGAAGTTTACGAAGTTGCACTGTCTGTGGGCTCAACCTTTTATTTGGTATGAAGCA	1830
DB	11946	TTTGAATTTTCAATGTATGACACTTGCCACATGTGCTCAACCAATTTATTAAGTAAAGCCA	11887
QY	1831	CACCTTTTACAGATGCTGATTAAGCAATTCACAGCACTTGACAGGCTAAGATATGCT	1890
DB	11886	CAATTCATCATGATGAGGGGAGATCTTTTGAACCACTCCAGAAAGTTCAACATATAGTCT	11827
QY	1891	ATGAGAGATTCCTCGCATGTGTATGATACACGCAATATTTGGAGATAGACTTCGCTCA	1950
DB	11826	CATTATGCTTATG-----ATATATAGATGACCTTGCATTTG	11791
QY	1951	GAAAGGGGACTCTGACAGCTGCTATATAGGCAAGTTGATGATATGCGTATGTTGGTGGG	2010
DB	11790	GAAAGGGGACTCTGCTCATTTCTATGTGAGATTTTGAAGATGTGGCTTTGA	11731
QY	2011	TTAGACATGAGGATTCACATATATAGGATTCAGCTATTTGTGAGATTGTGTTTGAAGAT	2070
DB	11730	ATTGATATATAGGTCTTCACTCATACAGAGACCCCAAAATTTCTAAGTTTATTTGACCAAC	11671
QY	2071	TC---AAATGCTGATGACATGATGATCTCCCTGAGCTATGCAAAATTTGTGGAACCTGG	2127
DB	11670	TCATGACATCACTGAGAAAGAAATCTTCTTCCAGAGGCTGTGCAAGCTTTTGGAGACTTGG	11611
QY	2128	TTGGCAGGGGTGTCTTTCTTAAGGTTCAAGAGACCAAAATATTAATACTCGGG	2187
DB	11610	CTTATCTTTAGAGTTTTTTCTTAAGAGCAGAGATCTCGGGGCACTGCAGTTCAACTTTGA	11551
QY	2188	GACTACTATGATGATCTATGTTTTTATGATTTCTTGAAAGATGAGGATGTTCAAGGT	2247

Db	11550	GATTACTACAGATGATCCAGAGTTTAAAGTACTGTAAGAAAGATGAGAGGTGTGTGCT	11493
Oy	2248	TCCTCTTACTGCTGCTGCAGACTATGSCAMAGATTGAGC-----C	2289
Db	11490	TCTCATTTGGCTGTGTGCTGTGCTATTTGAAAACCTTGGGTCTCAAGCTACAGCTACGCTGACCT	11433
Oy	2290	GAGCATGTGAAGCTAGTGTCTATGCAAGGCATGCAAGAAATTTCCTTCCGCTATPACA	2349
Db	11430	GGTACTGTGAATCAATATGCTATTCAAGCGCTTCACAAAGGTTTTCATATGATAGAACAG	11372
Oy	2350	GATGAAACCTCGGTGTAACCAAGATGTGCAAGACAGTGTTAGTGTAGATCCGTGT	2409
Db	11370	TTAGACAGGTGACGTAGAAAAATPCTAAAGATGSCCTTGGGGGATATCTTGAATAATTT	11311
Oy	2410	GGTACAAATGTAGCCGTGATGTGAGCCCTGG-----TGCTTTAT	2451
Db	11310	GACCAAGAAATATGACCTGCTCATGATTTGAGAAATGCGCCCTGAAAGATTAATCTGTCT	11255
Oy	2452	GCAGAGCTGTAGAACCTCTGTAAAACTTTGAAACTATATGATTAATGCAATTGAGCTGGG	2511
Db	11250	GGCCCACTGTTTGCACTGTTGGCAGTAATTTGGGCCCAAAATATTGGCCGTGAAGAGCCC	11199
Oy	2512	GTCACAGAGTACCGCTTGATGAAACTACGTGTGAATATGCGTCCGTCAATATGTTAAAG	2571
Db	11190	CTTTCCTATTTAGAGATGAGATGATCTGTGGCAGTGTGCTAATATGTGTCACCTTACT	11133
Oy	2572	GAGCAGATGTGAAGATCCTAGCTGCTGTGTGGCAATGGAATGCACTGTTCACTGTTCAAC	2631
Db	11130	GATGTGATGTAATGCTTTTCATT--TGCTAAGTGTGTAACAAACACTGGAATTTACG	11072
Oy	2632	CAGAAATATTTCTTAAAGCAGCTCATCTTTCAACGCAAGATATGTTCTTCTATG	2691
Db	11072	CTACGCTTACTGCGGTATCCTTGTGGCTGATATCAGATGTCTTATCTTATAGCT	11013
Oy	2692	GAATCTGATGTCCGCTACCATAGGTATGATTTAAATGATGCAATTTTCAATATCTGATTTG	2751
Db	11012	TTGCATTAATGTTTAAACAAAGCTATACATCAAAATTTTGTCTATACATACAAAG	10953
Oy	2752	CTCAAAATATGCTTGTGTTTGTGAGCTAAGAACATPATGTCCCACTTAATCATGTCCCAA	2811
Db	10952	CTACTATTAATTCGTATATGTAATATCATAGAGTTAACTTTGATGATGAATGTTGAAG	10893
Oy	2812	AGTGTACCAAGATTAACAAGTGTCTGAGTAAATTTCAATAATATGCTGCTG-----A	2866
Db	10892	ATCCTGCACTAATGAAGATCCAGTACATATTCCTGAATGATGGAAGCTGGCAGAG	10833
Oy	2867	ATTTTGTGATCAACTGTAGACAGAAATGTAAATTTCACTTCATCTTCATATTTCTGTTTGA	2926
Db	10832	ATATTTGTCAGAGTGCAGAGATCAAAATCTAAGGCTTTGGGACAGAAACATTCGGTGG	10773
Oy	2927	TAAAGTGAATTAAGATTGCGCTTAATGTGGCTTTGCCAATTTTCTTTC--CTTGATTT	2985
Db	10772	CATATTTGCAAGAGTGATTTTAAACCTTTGACTTTTCAATGCTTCTTACGTGGCCACTG	10713
Oy	2986	TTTTCTTTTGATTTAGGTGACGTCAAGCTGACGATTAAGAGCACTTCCAGATGGA	3045
Db	10712	CTCTTACTTAATCCTACGTTCTCTTAATAGAGATTAATTAAGATATATCTCTTAATAGGA	10653
Oy	3046	TGCTAGACGTGCAGAGATTAAGTATTCAGATGGCAGAAAGATTAACTCTGTGGCTTTGG	3105
Db	10652	TATATTTGTACCCCAATGTACTTTTCAAGTTTGTATTAACAATATCTTCT-----	10602
Oy	3106	GCCTGATCACCGGATGAATATGTTACAGAGGTGAGGAAATATATCAATTCATCA	3165
Db	10601	---TGTTGACTCACACATAGCTTAAATTTGCACTGAACATTTTTCAGAACTATTC	10545
Oy	3166	TTGTGTGAAAACGTGTGGAATGATTAATAGTCTGTGCTGTGTTGATTT--CTGTAT	3222
Db	10544	CACGTGATGATATCCCGGTGATATGGCGGTCAAAATGCTGTGATATTTTTCATGTTCT	10485
Oy	3223	TATATGTTTGTGATGGCGGAATCTGAATTTTGACTGACAGAGAGCTGAATCTGGCC	3282

Db 10484 TATAGTTCCTTGATGGCAACATGCTAAAGGTGTGSACTGACCGAGCGAGATTTAGC 10425
Qy 3283 AGCTTGGGTTGGTTTATGATATACACTGTGTGAACATATCTGTGACAGTGTGACAGTCT 3342
Db 10424 GTATGGGTGGTTTGGAGATATACATATCCGATGTGACATTTATGATATGATCTATCT 10365
Qy 3343 CAGCAGATGGAACCCGTCTGTGTGGAAGCACTGTGAGAGTCTGTGCTTGTCTTATCTG 3402
Db 10364 CCCATGATGTGCGACGAGCGACTGTGAGGCTTACGATTTGATGAGAGCGCCACCTTACTG 10305
Qy 3403 ATTGTGTTCAATCCGAAAACAAATGCTACTGATGTCAAGAACTTACAAACAGATACGAAG 3462
Db 10304 ATGTACTAGAGCCGAAACATATATTCATATGACAAATTAACCTACCCGATATGGA 10245
Qy 3463 TTTCTGTGTCGAAG---TCAGGGTGGAATAATCACTGAAGCTGTGTTCTTGCATCATPAT 3519
Db 10244 TGGCTCTTCCAGACTGAGAGGTGGAAAGATACGAAGAGAGCTCTCAAGTCTGAGA 10185
Qy 3520 ATACTCATATGTAGCATGTC 3539
Db 10184 ACAGTGTGACGTAGCTTGTCTC 10165

RESULT 8
AY19896 341 bp DNA linear PLN 24-JAN-2003
LOCUS Arabidopsis thaliana sequence flanking Ds3 end of Ds-EnhancerTrap
DEFINITION Insertion from line ET7222.
ACCESSION AY19896
VERSION AY19896.1 GI:27895850
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
AUTHORS 1 (bases 1 to 341)
May, B.P., Simorowski, J., Arroyo, J.-M., Vaughn, M.W., Shen, R.,
McCombie, W.R. and Martienssen, R.A.
TITLE Arabidopsis genomic sequences flanking Ds enhancer and gene traps
in transgenic lines
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 341)
AUTHORS May, B.P., Simorowski, J., Arroyo, J.-M., Vaughn, M.W., Shen, R.,
McCombie, W.R. and Martienssen, R.A.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-2002) Cold Spring Harbor Laboratory, 1 Bungtown
Road, Cold Spring Harbor, NY 11724, USA
COMMENT http://genetrap.cshl.org.
FEATURES
Location/Qualifiers
1..341
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
/clone="ET7222.Ds3.03.05.2002.jw20.341"
/ecotype="Landsberg"
/note="transgenic line ET7222"

ORIGIN

Query Match 8.3%; Score 304.4; DB 8; Length 341;
Best Local Similarity 99.4%; Pred. No. 3.3e-61;
Matches 316; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 497 ACCTGCGATTTGCTCTCCCATTTCAATATAGCGGATTACGACGGCGACGACAAAGC 556
Db 2 ACCTGCGATTTGCTCTCCCATTTCAATATAGCGGATTACGACGGCGACGACAAAGC 60
Qy 557 TCCGACGTAGCAACACCTCTACAACTATCTCTCCGCGACGAATGGCGCACCGTCTC 616
Db 61 TCCGACGTAGCAACACCTCTACAACTATCTCTCCGCGACGAATGGCGCACCGTCTC 120
Qy 617 TTCTCTCGAATTCAATTTCACTCCGATTTCTCTCTCTCTCTCTGCGCACCGCGCACCA 676

Db 121 TTCTCTCGAATTCAATTTCACTCCGATTTCTCTCTCTCTCTCTGCGCACCGCGCACCA 180
Qy 677 CCAACGGCACTCTGCTCTCTGCGACCACTATATTGATCGTCCGAACGGCACGTCCCA 736
Db 181 CCAACGGCACTCTGCTCTCTGCGACCACTATATTGATCGTCCGAACGGCACGTCCCA 240
Qy 737 TCCCATTTGATTTCTACACGATTTAGAGCTCAACACATTTCTTAAACGATGGAATCA 796
Db 241 TCCCATTTGATTTCTACACGATTTAGAGCTCAACACATTTCTTAAACGATGGAATCA 300
Qy 797 GAAGAGCATTCGAAGCTA 814
Db 301 GAAGAGCATTCGAAGCTA 318

RESULT 9
166494 7218 bp DNA linear PAT 28-DEC-1997
LOCUS Sequence 14 from patent US 5670367.
DEFINITION 166494
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 7218)
AUTHORS Dörner, F., Schefflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
Location/Qualifiers
1..7218
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 1.6%; Score 57.6; DB 6; Length 7218;
Best Local Similarity 3.9%; Pred. No. 0.019;
Matches 15; Conservative 218; Mismatches 147; Indels 0; Gaps 0;
Qy 510 TCTCTCCCATTCATATTATGCGATTACCGACGCGCGCAACAACTCCGACGTAGCCA 569
Db 1078 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1137
Qy 570 CAACACCTCTACAACTATCTGCTCCGCGCAAAATGGCGCACCGCTTCTCTCGACTT 629
Db 1138 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1197
Qy 630 CAATTTCACCTCGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 689
Db 1198 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1257
Qy 690 CGTCTCTGCGACCACTATATTGATCGTCCGGAACGCGACGTCGCCATTCCTGATTTT 749
Db 1258 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1317
Qy 750 CTACACGATTTAGAGCTCAACACATTTCTTAAACGATGAAATCAAGACATTCGA 809
Db 1318 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1377
Qy 810 AGCTAGGTTTCGAACCGCGCAATTCGTTTCAGAGAGAGCGCTTATACGCCGAG 869
Db 1378 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1437
Qy 870 ACAGATTCCTCAAGTCTT 889
Db 1438 CCAATTCCTTATCTCTT 1457
RESULT 10
AX083744 1141 bp DNA linear PAT 28-FEB-2001
LOCUS Sequence 22 from Patent WO011061.
DEFINITION AX083744
ACCESSION AX083744

VERSION	AX083744.1	GI:13185472
KEYWORDS		
SOURCE	synthetic construct	
ORGANISM	synthetic construct	
	other sequences; artificial sequences.	
REFERENCES	1	
AUTHORS	Kunst, J. and Clemens, S.	
TITLE	Regulation of embryonic transcription in plants	
JOURNAL	Patent: WO 011061-A 22 15-FEB-2001;	
	UNIVERSITY OF BRITISH COLUMBIA (CA)	
FEATURES	Location/Qualifiers	
source	1..1141	
	/organism="synthetic construct"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:32630"	
	1..1141	
	/note="consensus sequence of A.t., L.a., and B.n. PAB1 promoters"	
ORIGIN		
Query Match	1.5%, Score 54; DB 6; Length 1141;	
Best Local Similarity	10.0%; Pred. No. 0.12;	
Matches	79; Conservative 311; Mismatches 389; Indels 8; Gaps 2;	
QY	2293	TTAAGTGAAGATCCCTGTTGGTGAACAATGAGCCCGGTGAGTGGACCGTGGTCTTTATG 24523
DB	67	KYRRMYNNKSRWVGKWKYBCANNTSBRYYHARWMDMTAYABMTTNGKTGWRHR 126
QY	2453	CAGAAAGCTGAAGACCCCTGAA-ACCTTGAACATATGATTAAGCAATTCGAGCTGG 25111
DB	127	YMRBRABDYLDDHYYTANNAATWTCMDKDDKTRMMKGNNAATGMDDTKYHMMN 186
QY	2512	GTCTCAGAGTACGCTGATGAAACTAGTGAATGTCGGTGTGATATGTAAAG 25711
DB	187	NGCGTVMVYKTKDRDMSBKMYUQGBMYKMWISYDYUWVMDMDCKLVRRMVRIG 246
QY	2572	GAGCAAGTGAAGAGTCTTAGCTGCTGGTGGTGGCAATGAGCTGATTCACCTTCACG 26311
DB	247	RMRMYMAWBAHRRRRYNNGMTBMAHARMTNNNNNNNAKMKCRKAKWGNRAVNSTC 306
QY	2632	CAGAAATATTTCTTAAAGAAGCTCATCTTTCAAGCAAGATATGTTCTTATG 26911
DB	307	TTWKSATTKTATSCWANNCRAGDANKHKMKWMAKGVYNNNNNNNTWKARHBAW 366
QY	2692	GAATCTGATCGCTACATAGTATGATTAATGATGCAATTTTCATATATGCTATG 27511
DB	367	DMVHSAKKMNAADAHYSRKKMTBYRKATVNNNNNGTMMKRYAMWYKMDMBGTYN 426
QY	2752	CTCAAAATATGCTGTTTGTGAGCTTAAAGACATAGTTCSSCATTAATACATGCCAA 28111
DB	427	NNNNGGRTTYGWTYKMKMTYYKWKANNCKWRADHKTCTHNTTMMKTKYNNCWKS 486
QY	2812	AGTTGTACCAAGATTAACAAGTTCGTAGTAAATTTCACTAATATATGT-----GCTT 2864
DB	487	MTNGKSRBBAALVUTWMMRRYAHANNNNNDWDMWKACTYKYBVCSKMNNTAAATK 546
QY	2865	GAATTTTGTGACAACTGTGACAGACAAATGTAATTTCACTCTCAATTTCTGTAG 2924
DB	547	SSWMTTSRYTKTKTNNKRRMSDTRSMGRANAYARABHYGKNTKRMBSHTBHBAG 606
QY	2925	AATACAGTGAAGATTAAGAGTTCCTTAGTGTGGCTTTGTCCAACTTTTCTTCTGATT 2984
DB	607	AATYMMBMYBAKCHCKMYAKKYARAGGSNNNNNNNNNNNNNNNATCARDUYAA 666
QY	2985	TTTTTCTTTGATTTAGAGGTCAAGTCAAGCTGACATTCAGAAAGCACTTCCAGAAATG 3044
DB	667	RWYMANAKWYUYBAANNAAYTTAANNWGGCWNATDTRRTWKNNNNNNAGTKNNNN 726
QY	3045	ATGCTAGACATGACAGAGATATAGTATCAAGTGGCAGACAAATTAAGCTCTGGCTTTG 3104
DB	727	NARASAAKAYTAAAVDAKHKWRANMKARGNHDAABATTTDKRNGATTKTNTNN 786
QY	3105	GGCTGATGACCGCATGAAATGTTTACCAAGGTGAGGGAATTAATCTTCAATTCATCA 3164

[illegible]

```

IPTSKEIYTYKESKLDONDLKLSDEKETTKNISNDLELIKGEKKKNDPTFPI
SHNEKCKNYLPYDHKGDNDYNDKNNLLPCHDI CDNYNDKNNLLPCHDI CDN
SYNBEDLFTSLMSSTIPDKRGKIKMECKSIECVDDKNNLSYNDENNKNLLSYN
VOENEDKLYKNEQIKODIYNDNENDKLYKNEQIKODIYNDNENDKLYKNDOKI
IYNDNENDKLYKNDOKIKODIYNDNENDMLINSNPENTKYIKNIKENDKLLKNVKK
BEIDKREKSNICICKYIDKIDMKULIBEIKKKKKKNVENSNSYSDDPCFYEIYF
ICSNNIEGNNENIFDDTNINVENSNIDIMOSIFNNQIKYLDNLISQIFS"

repeat_region 1704..1740
/rpt_type=tandem
2309..2446
/rpt_type=tandem
2310..2443
/rpt_type=tandem
2313..2449
/rpt_type=tandem
2718..2893
/rpt_type=tandem
3065..3311
/rpt_type=tandem
3065..3273
/rpt_type=tandem
3776..3805
/rpt_type=tandem
4039..4094
/rpt_type=tandem
4260..4295
/rpt_type=tandem
4402..4439
/rpt_type=tandem
complement(<4519..>6088)
/locus_tag="PF14_0239"
complement(join(<4519..4588,4895..5573,5806..>6088))
/locus_tag="PF14_0239"
complement(join(4519..4588,4895..5573,5806..6088))
/locus_tag="PF14_0239"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAN36852.1"
/db_xref="GI:23497307"
/translation="MEEQNEIEKTVSESEINEEVEELPLFASDNDLVKESLRI
ILSLERKPIDHIIKNNKYIKTLISINSRCLLRLGLCLNLSAQLPELTERNLI
EILFDMMKBERKIEDNCIDYIMTISNLTCKSEYVYVLIDNDSNINIKEDNKFV
PLNKLLYFPFLPIKPSINKNLSDRYIVSHVLNIISSIKSYFFKVAFLNKISDGI
LNVFRPAIILPFIINLCINIEVIHPIFDHCYLPYVLSYLYNDVNIITSGSNNN
NNEINONTIHHIIMKKSILVSCVSKSRLLILFYLCNRDYSRNDIEMVTNGLIEG
SNREAIATAT"

repeat_region 4677..4714
/rpt_type=tandem
5600..5641
/rpt_type=tandem
5745..5791
/rpt_type=tandem
6108..6152
/rpt_type=tandem
6168..6222
/rpt_type=tandem
6220..6268
/rpt_type=tandem
6420..6481
/rpt_type=tandem
6563..6621
/rpt_type=tandem
6718..6780
/rpt_type=tandem
6719..6790
/rpt_type=tandem
7543..7568
/rpt_type=tandem
7761..7816
/rpt_type=tandem
7824..7849
/rpt_type=tandem
7871..7930
/rpt_type=tandem

```

```

gene /rpt_type=tandem
complement(<7953..>8438)
/locus_tag="PF14_0240"
mRNA complement(<7953..>8438)
/locus_tag="PF14_0240"
CDS complement(7953..8438)
/locus_tag="PF14_0240"
/codon_start=1
/product="ribosomal protein L21e, putative"
/protein_id="AAN36853.1"
/db_xref="GI:23497308"
/translation="MGKSRGSRSGTRYKSKKRRHGECTANKYLEKNYGVYD
VCDSTQCKMFPNRYHKGTCGIPHTRGGVVLNKRVRHIEQKVCVRIEHRKSC
NEDFLRKIKNAELIEAKLIKNNHINIKRTGSPKPAKIVPPSKIITIEPLPYEE
Y"

repeat_region 8467..8550
/rpt_type=tandem
8473..8539
/rpt_type=tandem
8477..8537
/rpt_type=tandem
8723..8844
/rpt_type=tandem
8723..8840
/rpt_type=tandem
8782..8844
/rpt_type=tandem
9019..9059
/rpt_type=tandem
9019..9058
/rpt_type=tandem
9133..9189
/rpt_type=tandem
9221..9254
/rpt_type=tandem
9604..9637
/rpt_type=tandem
9668..9723
/rpt_type=tandem
10036..10073
/rpt_type=tandem
10445..10472
/rpt_type=tandem
10511..10551
/rpt_type=tandem
complement(<10565..>11080)
/locus_tag="PF14_0241"
complement(<10565..>11080)
/locus_tag="PF14_0241"
/locus_tag="PF14_0241"
/codon_start=1
/product="basic transcription factor 3b, putative"
/protein_id="AAN36854.1"
/db_xref="GI:23497309"
/translation="MEKISPEILAAPAKLEKKNGNLRIGQGSARRKIKYKHNSI
SNEKINIIKIKGASFGVDIEI CVYRTGDTPLKPKPLCASLOSNTIYVGFKE
HKIDINKLFEGLKGNKYLDMNLLKIKINDPNIKNIIKESGDPKKEDEQANDVD
LVENPEVSKKE"

repeat_region 11429..11458
/rpt_type=tandem
11514..11584
/rpt_type=tandem
11514..11579
/rpt_type=tandem

Query Match 1.54; Score 54; DB 3; Length 250029;
Best Local Similarity 46.7%; Pred. No. 0.16;
Matches 171; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

OY 11 TAAGAGATACATATTAAACAATTGCTTGATTTCAACAAGATTTGGCTGCTATA 70
DB 218336 TAATAGATATGATGTGTAGATTATTATCTTGTCAGCAAGATTTATTATTGA 218395

```

QY		71	GGATTCATTGCGTCTGTGGTTGCCTTTACATTTCAAGTCATAAATAGTGGAATTTACAC	130
Db		218396	TTGAAAAGAGTGTTTTAATTTGCAATATGTCCTTTATTTGAAATAGCTTATATTTTTCC	218455
QY		131	ATTTCAGTTGGATGTGTAAGAAAAAGAGAGGAATGAGGGGTTTGTGGGTTAACTTT	190
Db		218456	AATCCTGATGATCATTTATTTGATGTAATTAATATATTAATTAACCTGCACATTTTCAATTT	218515
QY		191	AAAGTAGTCAGAATTAAGTCATTGGTTACTGTGCTCTATATGTTGTAATGAAGCA	250
Db		218516	ATATTAAATGGAATATGTTTTTTTTTGTCTTAAGATTTATTTTCTTTTAAATGTTACAT	218575
QY		251	ACTCCAAGGTTCTTAGTGGGAAATGATATTAATGACGATTAAACATCAAAGTCCGAG	310
Db		218576	ACATCGACAATGTGCAANAAGACACTATCTTCTTTTAACTTAATTTCTTAATATCT	218635
QY		311	GCGACTGTACATCATATGTTATGTTTTTATTTTTCAGTAGCTGTCGATGTTTTGAT	370
Db		218636	ATTCTGAGTAACCATTTTTTTTTTTTTTTTTTTTTTTTTTTTCATATTAATTTTGTAA	218695
QY		371	TTAATCT 376	
Db		218696	TTAAAT 218701	
RESULT 12				
AL713891		205796 bp	DNA	linear HTGS 30-MAY-2002
LOCUS		Mus musculus chromosome 11 clone RP23-435M14,	WORKING DRAFT	
DEFINITION		SEQUENCE.		
ACCESSION		AL713891.7	GI:21261921	
VERSION		HTGS_PHAUSE2; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.		
KEYWORDS		Mus musculus (house mouse)		
SOURCE		Mus musculus		
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE		Submitted (27-May-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgehire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk On May 29, 2002 this sequence version replaced gi:20067290.		
AUTHORS		Center: Genome Center		
JOURNAL		Center: Wellcome Trust Sanger Institute		
		Center code: SC		
		Web site: http://www.sanger.ac.uk		
		Contact: humquery@sanger.ac.uk		
		Project Information		
		Center project name: BM435M14		
		----- Summary Statistics -----		
		Assembly program: XCAP4; version 4.5		
		Chemistry: Dye-terminator; 15% of reads		
		Chemistry: Dye-terminator Big Dye; 84% of reads		
		Consensus quality: 201971 bases at least Q40		
		Consensus quality: 203010 bases at least Q30		
		Consensus quality: 203553 bases at least Q20		
		Insert size: 205796; sum-of-contigs		
		Insert size: 201568; 3.2% error; agarose-fp		
		Quality coverage: 5.53x in Q20 bases; sum-of-contigs Quality		
		coverage: 5.69x in Q20 bases; agarose-fp		
		----- NOTE: This is a 'working draft' sequence. It currently		
		* consists of 1 contigs. Gaps between the contigs		
		* are represented as runs of N. The order of the pieces		
		* is believed to be correct as given, however the sizes		
		* of the gaps between them are based on estimates that have		
		* provided by the submittor.		
		* This sequence will be replaced		
		* by the finished sequence as soon as it is available and		
		* the accession number will be preserved.		
		1 205796: contig of 205796 bp in length.		

[illegible]

repeat_region	/rpt_type=tandem	19987.720311
repeat_region	/rpt_type=tandem	19987.720311
repeat_region	/rpt_type=tandem	19987.720311
repeat_region	/rpt_type=tandem	20045.20265
repeat_region	/rpt_type=tandem	20175.703110
repeat_region	/rpt_type=tandem	20402.20538
repeat_region	/rpt_type=tandem	20446.20501
repeat_region	/rpt_type=tandem	20455.206110
repeat_region	/rpt_type=tandem	20605.20740
repeat_region	/rpt_type=tandem	20622.20825
repeat_region	/rpt_type=tandem	20786.20955
repeat_region	/rpt_type=tandem	20935.210310
repeat_region	/rpt_type=tandem	21037.21332
repeat_region	/rpt_type=tandem	21169.21467
repeat_region	/rpt_type=tandem	21529.21616
repeat_region	/rpt_type=tandem	21590.21994
repeat_region	/rpt_type=tandem	21625.21998
repeat_region	/rpt_type=tandem	21632.22441
repeat_region	/rpt_type=tandem	21632.22001
repeat_region	/rpt_type=tandem	21663.22441
repeat_region	/rpt_type=tandem	22127.22829
repeat_region	/rpt_type=tandem	22762.22842
repeat_region	/rpt_type=tandem	22875.23014
repeat_region	/rpt_type=tandem	22998.25249
repeat_region	/rpt_type=tandem	23670.23959
repeat_region	/rpt_type=tandem	23904.24125
repeat_region	/rpt_type=tandem	23989.24150
repeat_region	/rpt_type=tandem	24430.25018
repeat_region	/rpt_type=tandem	24527.24663
repeat_region	/rpt_type=tandem	24530.24751
repeat_region	/rpt_type=tandem	24617.24655
repeat_region	/rpt_type=tandem	25405.25446
repeat_region	/rpt_type=tandem	25464.25567
repeat_region	/rpt_type=tandem	25572.25714
repeat_region	/rpt_type=tandem	25763.26236
repeat_region	/rpt_type=tandem	26240.26333
repeat_region	/rpt_type=tandem	26382.26411
repeat_region	/rpt_type=tandem	

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 9, 2005, 16:58:33 ; Search time 2053 Seconds
(without alignments)
11072.555 Million cell updates/sec

Title: US-10-600-070-3

Perfect score: 3667

Sequence: 1 tggctgcattaaaggaggaat.....ctataacataaggctacaa 3667

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6046767 seqs, 3095530249 residues

Total number of hits satisfying chosen parameters: 12093534

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/prodata/1/pubpna/PC7_NEW_PUB.seq:*

3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*

5: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*

6: /cgn2_6/prodata/1/pubpna/PC7US_PUBCOMB.seq:*

7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*

14: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/prodata/1/pubpna/US10D_PUBCOMB.seq:*

17: /cgn2_6/prodata/1/pubpna/US10F_PUBCOMB.seq:*

18: /cgn2_6/prodata/1/pubpna/US10F_PUBCOMB.seq:*

19: /cgn2_6/prodata/1/pubpna/US10F_PUBCOMB.seq:*

20: /cgn2_6/prodata/1/pubpna/US10H_PUBCOMB.seq:*

21: /cgn2_6/prodata/1/pubpna/US10I_PUBCOMB.seq:*

22: /cgn2_6/prodata/1/pubpna/US10I_NEW_PUB.seq:*

23: /cgn2_6/prodata/1/pubpna/US11_PUBCOMB.seq:*

24: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*

25: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*

26: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3667	100.0	3667	19 US-10-600-070-3	Sequence 3, Appl1
2	3663.8	99.9	3667	19 US-10-600-070-10	Sequence 10, Appl1
3	1968.4	53.7	2679	20 US-10-739-930-227	Sequence 227, App
4	1931	52.7	2637	19 US-10-600-070-130	Sequence 130, App
5	1724	47.0	2406	19 US-10-600-070-1	Sequence 1, Appl1
6	1720.8	46.9	2406	19 US-10-600-070-9	Sequence 9, Appl1
7	1717.6	46.8	2406	19 US-10-600-070-128	Sequence 128, App

C	8	402.2	11.0	561	19	US-10-600-070-132	Sequence 132, App
	9	295.6	8.1	631	19	US-10-600-070-184	Sequence 184, App
	10	262.4	7.2	1146	18	US-10-424-599-129007	Sequence 129007, App
	11	208.8	5.7	660	19	US-10-600-070-135	Sequence 135, App
	12	194	5.3	2283	19	US-10-600-070-126	Sequence 126, App
	13	178.6	4.9	607	19	US-10-021-323-3562	Sequence 3562, App
	14	175	4.8	439	11	US-09-732-627A-2154	Sequence 2154, App
	15	161	4.4	545	19	US-10-600-070-175	Sequence 175, App
	16	159.4	4.3	552	19	US-10-021-323-3636	Sequence 3636, App
C	17	158	4.3	608	19	US-10-600-070-137	Sequence 137, App
	18	149.4	4.1	1411	20	US-10-425-115-81853	Sequence 81853, A
	19	146.2	4.0	2130	19	US-10-437-963-69932	Sequence 69932, A
	20	145.6	4.0	527	19	US-10-600-070-134	Sequence 134, App
	21	145.4	4.0	647	19	US-10-767-701-4069	Sequence 4069, App
	22	141	3.8	1039	18	US-10-424-599-35059	Sequence 35059, App
	23	139.8	3.8	1536	20	US-10-425-115-57452	Sequence 57452, A
	24	136.2	3.7	647	19	US-10-600-070-185	Sequence 185, App
	25	134.6	3.7	652	19	US-10-600-070-186	Sequence 186, App
C	26	132.2	3.6	563	19	US-10-600-070-178	Sequence 178, App
	27	130.2	3.6	446	19	US-10-600-070-151	Sequence 151, App
C	28	127.6	3.5	479	19	US-10-600-070-150	Sequence 150, App
	29	125	3.4	537	19	US-10-600-070-143	Sequence 143, App
	30	122.6	3.3	307	19	US-10-600-070-138	Sequence 138, App
	31	118.6	3.2	871	19	US-10-600-070-153	Sequence 153, App
C	32	117.8	3.2	1703	19	US-10-437-963-69933	Sequence 69933, A
	33	110.4	3.0	618	19	US-10-437-963-8187	Sequence 8187, App
	34	93.6	2.6	491	19	US-10-600-070-174	Sequence 174, App
	35	92.6	2.5	187	19	US-10-600-070-136	Sequence 136, App
	36	92	2.5	418	19	US-10-600-070-144	Sequence 144, App
	37	91.4	2.5	460	19	US-10-600-070-187	Sequence 187, App
C	38	90.4	2.5	420	19	US-10-600-070-176	Sequence 176, App
	39	89.4	2.4	604	19	US-10-600-070-147	Sequence 147, App
	40	89.4	2.4	653	19	US-10-600-070-148	Sequence 148, App
	41	89.4	2.4	1032	19	US-10-767-701-9474	Sequence 9474, App
	42	87.2	2.4	480	19	US-10-600-070-145	Sequence 145, App
	43	86.8	2.4	622	19	US-10-600-070-146	Sequence 146, App
	44	85.2	2.3	300	19	US-10-600-070-180	Sequence 180, App
	45	83	2.3	541	19	US-10-600-070-154	Sequence 154, App

ALIGNMENTS

RESULT 1

US-10-600-070-3

Sequence 3, Application US/10600070

Publication No. US20040139500A1

GENERAL INFORMATION:

APPLICANT: Osteoryoung, Katherine W.

APPLICANT: Viltha, Stanislaw

APPLICANT: Koksharova, Olga A.

TITLE OF INVENTION: Placitid Division and Related Genes and Proteins, and Methods of

TITLE OF INVENTION: Use

FILE REFERENCE: MSU-08153

CURRENT APPLICATION NUMBER: US/10/600,070

CURRENT FILING DATE: 2003-06-20

NUMBER OF SEQ ID NOS: 206

SOFTWARE: PatentIn version 3.2

SEQ ID NO 3

LENGTH: 3667

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-10-600-070-3

Query Match 100.0%; Score 3667; DB 19; Length 3667;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3667; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGTTCGATTAAAGAGATTAACATTTGCTGATTTCACAGATTTTG 60

Db 1 TGGTTCGATTAAAGAGATTAACATTTGCTGATTTCACAGATTTTG 60

QY 61 CTGGCTAAGAGATTCATGCTCTGTTTGTCTTTAATTAACATGTCATATAAGTTTCG 120
DB 61 CTGGGCTAATAGGATTCATGCTCTGTTTGTCTTTAATTAACATGTCATATAAGTTTCG 120
QY 121 AATTTTACATTTCAATTTGGAATGTTAAGAAAGAGGGGAATGAGGGGTTTGTGGG 180
DB 121 AATTTTACATTTCAATTTGGAATGTTAAGAAAGAGGGGAATGAGGGGTTTGTGGG 180
QY 181 TTTTAACTTTAAAGTAGTCAAGAAATTAAGTCATGTTTACTGTGCTATATGTGTAA 240
DB 181 TTTTAACTTTAAAGTAGTCAAGAAATTAAGTCATGTTTACTGTGCTATATGTGTAA 240
QY 241 AATGAGGCAACTCCAAAGGTTCTTAGGTGGAATAGATTATTTAGACGATTTAAATCAT 300
DB 241 AATGAGGCAACTCCAAAGGTTCTTAGGTGGAATAGATTATTTAGACGATTTAAATCAT 300
QY 301 AAAGTCGGTGGCGCATGTAACATGATGTTGTTTATTTTATTTTCAATGCTGGTAT 360
DB 301 AAAGTCGGTGGCGCATGTAACATGATGTTGTTTATTTTATTTTCAATGCTGGTAT 360
QY 361 GTTTTGTGATTTAACTTATCTCAAAATCAAAATTCATAACCTTAGACGACCAAA 420
DB 361 GTTTTGTGATTTAACTTATCTCAAAATCAAAATTCATAACCTTAGACGACCAAA 420
QY 421 CAGTCTCTCAATATGTATAACAGAACAAAGTTTGTAGTGGCTTAAAAAGACATCCC 480
DB 421 CAGTCTCTCAATATGTATAACAGAACAAAGTTTGTAGTGGCTTAAAAAGACATCCC 480
QY 481 ATGGAAGCTCTGATGCAAGTCGGCATGTCCTCCCATTTCCAAATTAATGCGCATTAACA 540
DB 481 ATGGAAGCTCTGATGCAAGTCGGCATGTCCTCCCATTTCCAAATTAATGCGCATTAACA 540
QY 541 CCGGCGACGACAAAGCTCCGACGTAGCCACAAACACTCTAACATCTGCTCCGACG 600
DB 541 CCGGCGACGACAAAGCTCCGACGTAGCCACAAACACTCTAACATCTGCTCCGACG 600
QY 601 AAATGGGCGCAGCGTCTTCTCTCGACTTCATTAACCTTCGATCTCTCTCTCC 660
DB 601 AAATGGGCGCAGCGTCTTCTCTCGACTTCATTAACCTTCGATCTCTCTCTCC 660
QY 661 TTGGGCGACCGGACCAACACCGGCACTCTGTCCTCTGCAACATCTTATTAATGTCGCC 720
DB 661 TTGGGCGACCGGACCAACACCGGCACTCTGTCCTCTGCAACATCTTATTAATGTCGCC 720
QY 721 GAAGCGCACTCCCATCCCATTTGATTTTCTACAGAGTATTAAGAGCTCAAAACATTTTC 780
DB 721 GAAGCGCACTCCCATCCCATTTGATTTTCTACAGAGTATTAAGAGCTCAAAACATTTTC 780
QY 781 TTAACCGATGGAATCAGAAAGCATTCGAAAGTGGGTTTCGAAACCGCGCAATTCGGT 840
DB 781 TTAACCGATGGAATCAGAAAGCATTCGAAAGTGGGTTTCGAAACCGCGCAATTCGGT 840
QY 841 TTTCAGGACGACCGCTTTAATCAGCCGGAACAGATTCTTCAAGTGTCTTCGAAACTCTG 900
DB 841 TTTCAGGACGACCGCTTTAATCAGCCGGAACAGATTCTTCAAGTGTCTTCGAAACTCTG 900
QY 901 TCTAATCTCGGCTCTAGAAAGAGTACATGAAGTCTTCTTGATGATGAAGAGCTACA 960
DB 901 TCTAATCTCGGCTCTAGAAAGAGTACATGAAGTCTTCTTGATGATGAAGAGCTACA 960
QY 961 GTCATCATGATGTTCTTGAGGATTAAGGTATTCGATTCGAAATATAAGTTTCTTC 1020
DB 961 GTCATCATGATGTTCTTGAGGATTAAGGTATTCGATTCGAAATATAAGTTTCTTC 1020
QY 1021 GTTTTAAATTTCAAGAAATGGAATAAGAGAACTTTTATCTAGTGAAGTCTCTGGGGC 1080
DB 1021 GTTTTAAATTTCAAGAAATGGAATAAGAGAACTTTTATCTAGTGAAGTCTCTGGGGC 1080
QY 1081 TCTCTGTGTATTTGCAAGAGGTGTGAGACTGAGATAGTCTTCTGGGTTGTGAGGCTCT 1140
DB 1081 TCTCTGTGTATTTGCAAGAGGTGTGAGACTGAGATAGTCTTCTGGGTTGTGAGGCTCT 1140
QY 1141 GCTTAAGAGAGGTGCTAAGTCGTTTAAAGCAAGATGTGGTTTATGTTATGGCGCTTGC 1200

DB 1141 GCTTAAGAGAGGTGCTAAGTCGTTTAAAGCAAGATGTGGTTTATGTTATGGCGCTTGC 1200
QY 1201 GTTTCTGATGTCTCGAGGAGTGTATGCAATGGATCCACCTGATTTTATTAATCTGCTTA 1260
DB 1201 GTTTCTGATGTCTCGAGGAGTGTATGCAATGGATCCACCTGATTTTATTAATCTGCTTA 1260
QY 1261 TGAATTTGTGAAGAACCTTTGAAGCTTTTACAGATAGTTGAAGCTTGTGTAATTTTG 1320
DB 1261 TGAATTTGTGAAGAACCTTTGAAGCTTTTACAGATAGTTGAAGCTTGTGTAATTTTG 1320
QY 1321 ACGAGCGTTGGCTTTATTAAGAACTTTCTTGATTTGATCTTTGTATTTAGTCTGTGTA 1380
DB 1321 ACGAGCGTTGGCTTTATTAAGAACTTTCTTGATTTGATCTTTGTATTTAGTCTGTGTA 1380
QY 1381 GGAAGAGGAGCAAGTAGCTTGACCGGATTTACGGACCAATTAAGTAGAGCTTTTGA 1440
DB 1381 GGAAGAGGAGCAAGTAGCTTGACCGGATTTACGGACCAATTAAGTAGAGCTTTTGA 1440
QY 1441 AGAGATCACTCCGCGTTATGCTTGAAGCTACTTGGCTTACCGCTTGATGATTAACGC 1500
DB 1441 AGAGATCACTCCGCGTTATGCTTGAAGCTACTTGGCTTACCGCTTGATGATTAACGC 1500
QY 1501 TCGGAAAAGACTAATGTGTTTAAAGCGGTGTGCGGAATTTTGTGTCTGTGAGGAGG 1560
DB 1501 TCGGAAAAGACTAATGTGTTTAAAGCGGTGTGCGGAATTTTGTGTCTGTGAGGAGG 1560
QY 1561 TGGAGCATCAGCTCTTGTGGGGGTTTGAACCGGTGAAGATTATGAATGAGCGTTTTT 1620
DB 1561 TGGAGCATCAGCTCTTGTGGGGGTTTGAACCGGTGAAGATTATGAATGAGCGTTTTT 1620
QY 1621 ACGAATGACAGCTGTGAGCAGGATTAACGTTTAAATACCTTTTAAATTTCTTTAGCA 1680
DB 1621 ACGAATGACAGCTGTGAGCAGGATTAACGTTTAAATACCTTTTAAATTTCTTTAGCA 1680
QY 1681 TGAATTAATCTTATAGTTTCTCATTTTATATATATGTTGTGTAGTGAATCTTTTGTGA 1740
DB 1681 TGAATTAATCTTATAGTTTCTCATTTTATATATATGTTGTGTAGTGAATCTTTTGTGA 1740
QY 1741 GCTAACCCCAAGCAATATTCAGAGAGTCAATTGAAGTTTACAGAGTGTGACTTGCCTT 1800
DB 1741 GCTAACCCCAAGCAATATTCAGAGAGTCAATTGAAGTTTACAGAGTGTGACTTGCCTT 1800
QY 1801 GTGGCTCAAGCTTTTATTTGTGTAAGAACCAACCTTTTACAGATGTGATTAAGCAATTC 1860
DB 1801 GTGGCTCAAGCTTTTATTTGTGTAAGAACCAACCTTTTACAGATGTGATTAAGCAATTC 1860
QY 1861 CAGCAACTTCAAGCAGCTTAAGGTAATGCTATGGAATTCCTGCAATGTTATGATTAACA 1920
DB 1861 CAGCAACTTCAAGCAGCTTAAGGTAATGCTATGGAATTCCTGCAATGTTATGATTAACA 1920
QY 1921 CGGAATTAATTTGGAGATAGACTTCGCTTGAAGAAAGGGAATCTGTGCACTGTTATAGGC 1980
DB 1921 CGGAATTAATTTGGAGATAGACTTCGCTTGAAGAAAGGGAATCTGTGCACTGTTATAGGC 1980
QY 1981 AAAGTTGATGAATGCCGTAATGTGTTGGGCTTTAGACAGTGAAGATTACAAATATAGGAAT 2040
DB 1981 AAAGTTGATGAATGCCGTAATGTGTTGGGCTTTAGACAGTGAAGATTACAAATATAGGAAT 2040
QY 2041 CCAAGCTATTTGTGAAGTTTGTTTTGAAGAAATTCAAATCGTGAATACAAATGATATCCCT 2100
DB 2041 CCAAGCTATTTGTGAAGTTTGTTTTGAAGAAATTCAAATCGTGAATACAAATGATATCCCT 2100
QY 2101 GGACTATGCAAAATTTGTGGAACCTGCTGGGAGGAGGTTCTTCTCTAGGTTCAAGAGAC 2160
DB 2101 GGACTATGCAAAATTTGTGGAACCTGCTGGGAGGAGGTTCTTCTCTAGGTTCAAGAGAC 2160
QY 2161 ACCAAAGATTAATAATTTTAACTCGGGGACTACTATGATGATCTATGTTTGAATTAC 2220
DB 2161 ACCAAAGATTAATAATTTTAACTCGGGGACTACTATGATGATCTATGTTTGAATTAC 2220
QY 2221 TTGAAAAGAGTGAAGTATGTTCAAGGTTCTCTTTAGCTGTGCTGCACTATGCGCAAGG 2280

Db	2221	TTGGAAAAGTGGAGTATGTTCAAGGTTCTCTCTTACGTGCTGCAACATATGCAAGG	2280
Oy	2281	ATTGGAGCCGAGCATGTGAAAGCTATGTCTATGCAAGCATCTGCAGAAAGTTTTCTTCC	2340
Db	2281	ATTGGAGCCGAGCATGTGAAAGCTATGTCTATGCAAGCATCTGCAGAAAGTTTTCTTCC	2340
Oy	2341	CGCTATACATATAGAAACTCGGCTGTAACCCAGATGTGCAAGAGACAGTCTTATAGTGA	2400
Db	2341	CGCTATACATATAGAAACTCGGCTGTAACCCAGATGTGCAAGAGACAGTCTTATAGTGA	2400
Oy	2401	GATCCGTGTGGTAAACAATGTATAGGCCCGGTATAGGTGAGCCCTGGTGTCTTTATATGCAGAAAGCT	2460
Db	2401	GATCCGTGTGGTAAACAATGTATAGGCCCGGTATAGGTGAGCCCTGGTGTCTTTATATGCAGAAAGCT	2460
Oy	2461	GTAAGACCCCTCTGAATACTTTGAAACTATATGATTAATGCAATTCGAGCTGGGGTCTCAGAG	2520
Db	2461	GTAAGACCCCTCTGAATACTTTGAAACTATATGATTAATGCAATTCGAGCTGGGGTCTCAGAG	2520
Oy	2521	AGTACCGTTGATGAAACTATCTGTGAAATGTCCGTGTGTATATGTTAAAGAGACAGT	2580
Db	2521	AGTACCGTTGATGAAACTATCTGTGAAATGTCCGTGTGTATATGTTAAAGAGACAGT	2580
Oy	2581	GTTGAAGATCCATAGCGCGTGTGTGGCAATTGGAATTCGATTTACATCTTCAGCCAGCAAGAT	2640
Db	2581	GTTGAAGATCCATAGCGCGTGTGTGGCAATTGGAATTCGATTTACATCTTCAGCCAGCAAGAT	2640
Oy	2641	TTTCTTAAACACAGCTCATCTTTTCAACGCAAGATATGTTTCTTCTATGAAATCTGAT	2700
Db	2641	TTTCTTAAACACAGCTCATCTTTTCAACGCAAGATATGTTTCTTCTATGAAATCTGAT	2700
Oy	2701	GTCGCTACATATAGTATATTAATGATGCAATTTCAATATATCTGCAATGCTCAAAATA	2760
Db	2701	GTCGCTACATATAGTATATTAATGATGCAATTTCAATATATCTGCAATGCTCAAAATA	2760
Oy	2761	TGCTTTGTTTTGTAGCTTAAGAACATATGTTCCCATTTAATACATGTCACCAAAAGTTGTACC	2820
Db	2761	TGCTTTGTTTTGTAGCTTAAGAACATATGTTCCCATTTAATACATGTCACCAAAAGTTGTACC	2820
Oy	2821	AAGATTTAACAAAGTTGCTGATGTAATTTCACTAATATATGCTGCTGAATTTTTGATCAAA	2880
Db	2821	AAGATTTAACAAAGTTGCTGATGTAATTTCACTAATATATGCTGCTGAATTTTTGATCAAA	2880
Oy	2881	CTGTATGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTTAAATATACGTAGAAATTAG	2940
Db	2881	CTGTATGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTTAAATATACGTAGAAATTAG	2940
Oy	2941	AGATTTGCTTATGTGTGCGCTTGTGCCAATTTTCTTCCCTGATATTTTTTCTTTGCAATTT	3000
Db	2941	AGATTTGCTTATGTGTGCGCTTGTGCCAATTTTCTTCCCTGATATTTTTTCTTTGCAATTT	3000
Oy	3001	AGGGTCAGTCAAGGCTGACGATTCAGAGACACTTCCAGAAATGATGCTAGAGACTGAGCA	3060
Db	3001	AGGGTCAGTCAAGGCTGACGATTCAGAGACACTTCCAGAAATGATGCTAGAGACTGAGCA	3060
Oy	3061	GAATATATGATTCGAAGTGGCAGAAAGATTAAAGTCTCTGACTTTTGGGCCCTGATACCCGAT	3120
Db	3061	GAATATATGATTCGAAGTGGCAGAAAGATTAAAGTCTCTGACTTTTGGGCCCTGATACCCGAT	3120
Oy	3121	AGAAATGTTACCAAGGTGAGGGAAATTAATCTACATTCATCAATGTGTGAAAACCTGT	3180
Db	3121	AGAAATGTTACCAAGGTGAGGGAAATTAATCTACATTCATCAATGTGTGAAAACCTGT	3180
Oy	3181	TGGACATGATTAATAGTCTGTGTGCGCTGTGTTGATCTGTGTTATTTTAAAGTCTTTGATGGGC	3240
Db	3181	TGGACATGATTAATAGTCTGTGTGCGCTGTGTTGATCTGTGTTATTTTAAAGTCTTTGATGGGC	3240
Oy	3241	GAATGCTGAAGATTTTGACTGACAGAGCAGCTGAATCTGCAGCTTGGGTTGTATATG	3300
Db	3241	GAATGCTGAAGATTTTGACTGACAGAGCAGCTGAATCTGCAGCTTGGGTTGTATATG	3300
Oy	3301	ATTATATACCTGTTGAAACTATCTGTGTAACATGTGACAGTCTCAGACAGATGAAACCCGTG	3360
Db	3301	ATTATATACCTGTTGAAACTATCTGTGTAACATGTGTACAGTCTCAGACAGATGAAACCCGTG	3360

Qy		3631	CTGCGTGAAGCAACTCTGGAGAGTCGTGCTGCTATCTGATTGGTTGCATCCAGAAA	3420
Dd		3631	CTGCGTGAAGCAACTCTGGAGAGTCGTGCTGCTATCTGATTGGTTGCATCCAGAAA	3420
Qy		3421	ACAATGCTACTGATNTCAGAACCTTACACAAAGATACGAAGTTTTCTGCTCCAAGTCAG	3480
Dd		3421	ACAATGCTACTGATNTCAGAACCTTACACAAAGATACGAAGTTTTCTGCTCCAAGTCAG	3480
Qy		3481	GGTGGAATACTCATGAGGCTCTGTTCTTGCAATCATTAATTAATCTATGTAGCATGTCT	3540
Dd		3481	GGTGGAATACTCATGAGGCTCTGTTCTTGCAATCATTAATTAATCTATGTAGCATGTCT	3540
Qy		3541	GAGCTTGGAGAATCTCTTTGTTCTGTAATTCCTCTCTAAGTTAGTTGTTAAATGA	3600
Dd		3541	GAGCTTGGAGAATCTCTTTGTTCTGTAATTCCTCTCTAAGTTAGTTGTTAAATGA	3600
Qy		3601	ACACAAAAAATTAAAGTTCCTTGCGCACACCCTTTTCCTTGATCTAAACTTAACATAGG	3660
Dd		3601	ACACAAAAAATTAAAGTTCCTTGCGCACACCCTTTTCCTTGATCTAAACTTAACATAGG	3660
Qy		3661	GCTACAA 3667	
Dd		3661	GCTACAA 3667	

RESULT 2
US-10-600-070-10

```

; Sequence 10, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteeryoung, Katherine W.
; APPLICANT: Vichra, Stanislav
; APPLICANT: Koksharova, Olga A.
; TITLE OF INVENTION: Plasmid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600, 070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 3667
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-10-600-070-10
    
```

Query Match 99.9%; Score 3663.8; DB 19; Length 3667;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3665; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Db 301 AAAGTCGGCGCATGTAAATCATAGATGTTTTTATTTTTCAGTAGCGTGAGAT 360
Qy 361 GTTTTGTATTTAACTTATCTACTCAAAATCAAAATTCATAAACCCTAGACGACAAA 420
Db 361 GTTTTGTATTTAACTTATCTACTCAAAATCAAAATTCATAAACCCTAGACGACAAA 420
Qy 421 CAGTCTCTTCAATATGTAAACAGAACAAAGTTTTTGTAGAGCTTAAAAAGACACTCC 480
Db 421 CAGTCTCTTCAATATGTAAACAGAACAAAGTTTTTGTAGAGCTTAAAAAGACACTCC 480
Qy 481 ATGGAAGCTCTGAGTCAAGTCGGGATGTGTCTCTCCCATTCGAATTATGCGCATTAACA 540
Db 481 ATGGAAGCTCTGAGTCAAGTCGGGATGTGTCTCTCCCATTCGAATTATGCGCATTAACA 540
Qy 541 CCGGCGACGACAAAGCTCCGACGAGGACAAACACTCTACACATCTGCTCCGCGACG 600
Db 541 CCGGCGACGACAAAGCTCCGACGAGGACAAACACTCTACACATCTGCTCCGCGACG 600
Qy 601 AAATGGGCGACGCTCTTCTCTCCGACTTCAAATTTCACTCCGATTCCTCTCTCTCC 660
Db 601 AAATGGGCGACGCTCTTCTCTCCGACTTCAAATTTCACTCCGATTCCTCTCTCTCC 660
Qy 661 TTGGCCACCGGACACACACGCGCACTGTGTCTCTGCGACGATCTATTTGATCGTCC 720
Db 661 TTGGCCACCGGACACACACGCGCACTGTGTCTCTGCGACGATCTATTTGATCGTCC 720
Qy 721 GAAAGCGACGCTCCCATCCCATTTGATTTTCTACAGAGTATAGAGGCTCAAAACATTTG 780
Db 721 GAAAGCGACGCTCCCATCCCATTTGATTTTCTACAGAGTATAGAGGCTCAAAACATTTG 780
Qy 781 TTAAACGATGAATCAGAAAGACATTCGAACTAGGGTTTGGAAACGCGCAATTCGAT 840
Db 781 TTAAACGATGAATCAGAAAGACATTCGAACTAGGGTTTGGAAACGCGCAATTCGAT 840
Qy 841 TTCAAGCAGACGCTTTATATCAGCCGAGACAGATTTCTTCAAGTGTCTTGGAATCTGT 900
Db 841 TTCAAGCAGACGCTTTATATCAGCCGAGACAGATTTCTTCAAGTGTCTTGGAATCTGT 900
Qy 901 TCTAATCTCGGTCTAAGAGAGATCAATGAAGATCTTCTTGAATGAAGAGATCTACA 960
Db 901 TCTAATCTCGGTCTAAGAGAGATCAATGAAGATCTTCTTGAATGAAGAGATCTACA 960
Qy 961 GTGATCACTGATGTTCTTGGGATTAAGGTAAATTCGATTCGAAATATAAGTTCTTC 1020
Db 961 GTGATCACTGATGTTCTTGGGATTAAGGTAAATTCGATTCGAAATATAAGTTCTTC 1020
Qy 1021 GTTTTAAATTTCAATGAATTTGAATTAAGAAAGAACTTTTATCTAGTGAAGTCTGAGGC 1080
Db 1021 GTTTTAAATTTCAATGAATTTGAATTAAGAAAGAACTTTTATCTAGTGAAGTCTGAGGC 1080
Qy 1081 TCTCTGTATTTGCAAGAAAGGTGGTGAAGCTGAGATAGTTCTTGGGTTGTGAGGCTCT 1140
Db 1081 TCTCTGTATTTGCAAGAAAGGTGGTGAAGCTGAGATAGTTCTTGGGTTGTGAGGCTCT 1140
Qy 1141 GCTTAAGAGAGGTGGCTTAAGTGTTTAAGCAAGATGTGGTTTATAGTATGCGGCTTC 1200
Db 1141 GCTTAAGAGAGGTGGCTTAAGTGTTTAAGCAAGATGTGGTTTATAGTATGCGGCTTC 1200
Qy 1201 GTTTTCTGATGTCTCGAGGAGTGTATGGAATTCGACCTGATTTTATATCTGTTA 1260
Db 1201 GTTTTCTGATGTCTCGAGGAGTGTATGGAATTCGACCTGATTTTATATCTGTTA 1260
Qy 1261 TGAATTTGTGAGAACTTTGAAGCTTTTACAGGTAGTTGAATTTGCTTTGGTATTTG 1320
Db 1261 TGAATTTGTGAGAACTTTGAAGCTTTTACAGGTAGTTGAATTTGCTTTGGTATTTG 1320
Qy 1321 ACGAGCGTGGCTTATTAAGAACTTTCTGATTTGATTAATCTTTGATTTGATGCTTGTGA 1380
Db 1321 ACGAGCGTGGCTTATTAAGAACTTTCTGATTTGATTAATCTTTGATTTGATGCTTGTGA 1380
Qy 1381 GGAAGAAAGAGCAAGTAGCTTGACCGGATTTACGTGACAAATTTGATGAGACTTTGGA 1440
Db 1381 GGAAGAAAGAGCAAGTAGCTTGACCGGATTTACGTGACAAATTTGATGAGACTTTGGA 1440

Db 1381 GGAAGAAAGAGCAAGTAGCTTGACCGGATTTACGTGACAAATTTGATGAGACTTTGGA 1440
Qy 1441 AGAGATCACTCCGCGTATGTCTTGAAGTACTTGGCTTACCGCTTGGTATGATTAACGC 1500
Db 1441 AGAGATCACTCCGCGTATGTCTTGAAGTACTTGGCTTACCGCTTGGTATGATTAACGC 1500
Qy 1501 TCGGAAAAGCTAAATGTGTTTAAAGCGGTGTGCGGAATATTTTGTGTCTGTGAGAGAG 1560
Db 1501 TCGGAAAAGCTAAATGTGTTTAAAGCGGTGTGCGGAATATTTTGTGTCTGTGAGAGAG 1560
Qy 1561 TGGAGCATCAGCTCTTGTGGGGGTTTGAACCGGAGAGAGTTTATGAATGAGCGTTTTC 1620
Db 1561 TGGAGCATCAGCTCTTGTGGGGGTTTGAACCGGAGAGAGTTTATGAATGAGCGTTTTC 1620
Qy 1621 ACGAATGACAGCTGCTGAGAGAGTATACAGTTTATAGTATTTTATTTTATTTCTTGA 1680
Db 1621 ACGAATGACAGCTGCTGAGAGAGTATACAGTTTATAGTATTTTATTTTATTTCTTGA 1680
Qy 1681 TGAATTAATCTTATAGTTTCTCATTTTATATGATGTGTGTAGTATGATCTTTTGTGA 1740
Db 1681 TGAATTAATCTTATAGTTTCTCATTTTATATGATGTGTGTAGTATGATCTTTTGTGA 1740
Qy 1741 GCTAACCCGAGCAATATTCAGAGAGTCAATTTGAAGTTTACGAAGTTGACCTTGTCT 1800
Db 1741 GCTAACCCGAGCAATATTCAGAGAGTCAATTTGAAGTTTACGAAGTTGACCTTGTCT 1800
Qy 1801 GTGGCTCAAGCTTTTATTTGTTAAGAAAGCAACACTTTTACAGATGCTGATTAAGCAATTC 1860
Db 1801 GTGGCTCAAGCTTTTATTTGTTAAGAAAGCAACACTTTTACAGATGCTGATTAAGCAATTC 1860
Qy 1861 CAGCACTTCAAGAGGCTTAAGGTATATGCTATAGAGATTCCTGATGTGTATGATTAACA 1920
Db 1861 CAGCACTTCAAGAGGCTTAAGGTATATGCTATAGAGATTCCTGATGTGTATGATTAACA 1920
Qy 1921 CGGATTAATTTGGAGATATAGCTTCCGCTTGAAGAGGGACTCTGTGACAGCTTATAGGC 1980
Db 1921 CGGATTAATTTGGAGATATAGCTTCCGCTTGAAGAGGGACTCTGTGACAGCTTATAGGC 1980
Qy 1981 AAAGTGAATGAATGCCGATATGTGTGGGCTTGAACAGTGAAGATTCACAAATATAGGAAT 2040
Db 1981 AAAGTGAATGAATGCCGATATGTGTGGGCTTGAACAGTGAAGATTCACAAATATAGGAAT 2040
Qy 2041 CCAAGCTATGTGAGATTTGTTTGAAGAAATTCAAATCGTGTATACAAATGATATCTCCCT 2100
Db 2041 CCAAGCTATGTGAGATTTGTTTGAAGAAATTCAAATCGTGTATACAAATGATATCTCCCT 2100
Qy 2101 GGAATATGCAAAATTTGTTGAAGAACTGTGGTGGAGGGGTTGCTTCTCTAGGTTCAAGAG 2160
Db 2101 GGAATATGCAAAATTTGTTGAAGAACTGTGGTGGAGGGGTTGCTTCTCTAGGTTCAAGAG 2160
Qy 2161 ACCAAAGATTAATAATTTTAAACTCGGGGACTACTATGATGATCTATGTTTTGATTAAC 2220
Db 2161 ACCAAAGATTAATAATTTTAAACTCGGGGACTACTATGATGATCTATGTTTTGATTAAC 2220
Qy 2221 TTGAAAAGAGTGAAGGTATTCAGGGTTCTCTTATAGCTGCTGAGCTATAGGCAAGG 2280
Db 2221 TTGAAAAGAGTGAAGGTATTCAGGGTTCTCTTATAGCTGCTGAGCTATAGGCAAGG 2280
Qy 2281 ATTGAGCGGAGATGGAAGGTAGTGTATGCTATGACGCACTGAGAAAGTTTTTCTCTCC 2340
Db 2281 ATTGAGCGGAGATGGAAGGTAGTGTATGCTATGACGCACTGAGAAAGTTTTTCTCTCC 2340
Qy 2341 CGCTATACAGATTAAGAACTTCGCGTGAACCCAGAGATGTGCAAGAGACAGTATTAGTGA 2400
Db 2341 CGCTATACAGATTAAGAACTTCGCGTGAACCCAGAGATGTGCAAGAGACAGTATTAGTGA 2400
Qy 2401 GATCTGTGTGTAACAATGTAAGCCGATGATGTGAGCTGTGTCTTTATTTGCAAGAGCT 2460
Db 2401 GATCTGTGTGTAACAATGTAAGCCGATGATGTGAGCTGTGTCTTTATTTGCAAGAGCT 2460
Qy 2461 GTAAGACCCCTGGAAGAACTTTGAACATTAATGATTAAGCAATTCGAGCTGGGGTCTCAGAG 2520
Db 2461 GTAAGACCCCTGGAAGAACTTTGAACATTAATGATTAAGCAATTCGAGCTGGGGTCTCAGAG 2520
```

QY	2521	AGTAGCGTTGATGTAACATACCTGTTGAAATGTCGTTGCTGATATAGTTAAAGAGGCAAGT	2580
Db	2521	AGTAGCGTTGATGTAACATACCTGTTGAAATGTCGTTGCTGATATAGTTAAAGAGGCAAGT	2580
QY	2581	GTCGAAGATCCTAGCTGCTGTGTGGCAATTGAGCTGATTTCACTGTTCAAGCAGAAATAT	2640
Db	2581	GTCGAAGATCCTAGCTGCTGTGTGGCAATTGAGCTGATTTCACTGTTCAAGCAGAAATAT	2640
QY	2641	TTTCTTAAAGCAGCTCATCTTTTCAACGCAAGATATGTTTCTTCTATGAAATCTGAT	2700
Db	2641	TTTCTTAAAGCAGCTCATCTTTTCAACGCAAGATATGTTTCTTCTATGAAATCTGAT	2700
QY	2701	GTCCCTACCAATGATATTAATTAATGAAATTTTCAATATATGCTGAATTTTGTATCAAA	2760
Db	2701	GTCCCTACCAATGATATTAATTAATGAAATTTTCAATATATGCTGAATTTTGTATCAAA	2760
QY	2761	TGCTTGTTTTGTAGCTAAGAACATAGTCCCATTAATACATGTCCCAAAAGTTGTACC	2820
Db	2761	TGCTTGTTTTGTAGCTAAGAACATAGTCCCATTAATACATGTCCCAAAAGTTGTACC	2820
QY	2881	AAAGATTAAACAAGTTGCTGAGTAAATTTCACTAATTAATGCTGAAATTTTGTATCAAA	2880
Db	2881	AAAGATTAAACAAGTTGCTGAGTAAATTTCACTAATTAATGCTGAAATTTTGTATCAAA	2880
QY	2881	CTGTAGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTAAGATTAACGTAGATTAAG	2940
Db	2881	CTGTAGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTAAGATTAACGTAGATTAAG	2940
QY	2941	AGATTGCTTAATGTCGTGCTTGTCCAACTTTCTTCCCTGAATTTTTCTTTGCATTT	3000
Db	2941	AGATTGCTTAATGTCGTGCTTGTCCAACTTTCTTCCCTGAATTTTTCTTTGCATTT	3000
QY	3001	AGGGTCAGTCAGACCTGACGATTCAGAGACACTTCCAGATGAGTGTAGAGCTGACAGA	3060
Db	3001	AGGGTCAGTCAGACCTGACGATTCAGAGACACTTCCAGATGAGTGTAGAGCTGACAGA	3060
QY	3061	GAATATAGTATCCAAATGTCAGAAAGATTAAAGTCTCTGGCTTTTGCGCTGTACCCGAT	3120
Db	3061	GAATATAGTATCCAAATGTCAGAAAGATTAAAGTCTCTGGCTTTTGCGCTGTACCCGAT	3120
QY	3121	AGAAATGTTACCAAGGTGAGGGGAATTAATCTTACAATTCATTAATGTGTGAAACCTGT	3180
Db	3121	AGAAATGTTACCAAGGTGAGGGGAATTAATCTTACAATTCATTAATGTGTGAAACCTGT	3180
QY	3181	TGGAATGATTAATAGTCTGTGCTCTGTTGATTCGTGTAATTTTAATGATTTTGGATGGGC	3240
Db	3181	TGGAATGATTAATAGTCTGTGCTCTGTTGATTCGTGTAATTTTAATGATTTTGGATGGGC	3240
QY	3241	GAATGCTGAAGATTTTGAATGACAGAGCAGCTGAACTTCGCGACGCTTGGTGTGTTATG	3300
Db	3241	GAATGCTGAAGATTTTGAATGACAGAGCAGCTGAACTTCGCGACGCTTGGTGTGTTATG	3300
QY	3301	ATTATATACCTGTTGAACTTATCTGTTGACAGTGTGACGCTCAGCAGATGGAACCCGTG	3360
Db	3301	ATTATATACCTGTTGAACTTATCTGTTGACAGTGTGACGCTCAGCAGATGGAACCCGTG	3360
QY	3361	CTCTGCTGGAAGCAACTCTGGAAGAGTGTGCTGTCTATCTGAATTTGGTTCATCAAGAA	3420
Db	3361	CTCTGCTGGAAGCAACTCTGGAAGAGTGTGCTGTCTATCTGAATTTGGTTCATCAAGAA	3420
QY	3421	ACAAATGCTACATGATCTGAAACTTACACAAACAAAGTATTCGTGTCAAGTCTAG	3480
Db	3421	ACAAATGCTACATGATCTGAAACTTACACAAACAAAGTATTCGTGTCAAGTCTAG	3480
QY	3481	GGTGGAATATCACTGAAGGCTCTGTTCTTGATCAATAATATCAATATGTAGCATGCT	3540
Db	3481	GGTGGAATATCACTGAAGGCTCTGTTCTTGATCAATAATATCAATATGTAGCATGCT	3540
QY	3541	GAGCTTGGGAATTTCTTTTGTCTGTAAATTTCTCTCTAAGTTAGTTATTAATGA	3600
Db	3541	GAGCTTGGGAATTTCTTTTGTCTGTAAATTTCTCTCTAAGTTAGTTATTAATGA	3600

Qy	3601	ACACAAAAAATTAAAGCTCTTGGCAGACCCCTTCTCTGATCTAAACATTAACATTAAGG	3660
Db	3601	ACACAAAAAATTAAAGCTCTTGGCAGACCCCTTCTCTGATCTAAACATTAACATTAAGG	3660
Qy	3661	GCTACAA	3667
Db	3661	GCTACAA	3667
RESULT 3			
US-10-739-930-227			
; Sequence 227, Application US/10739930			
; Publication No. US20040216190A1			
; GENERAL INFORMATION:			
; APPLICANT: Kovalic, David K.			
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH			
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT			
; FILE REFERENCE: 38-21(53377)B			
; CURRENT APPLICATION NUMBER: US/10/739,930			
; CURRENT FILING DATE: 2003-12-18			
; NUMBER OF SEQ ID NOS: 11088			
; SEQ ID NO 227			
; LENGTH: 2679			
; TYPE: DNA			
; ORGANISM: Arabidopsis thaliana			
; FEATURES:			
; OTHER INFORMATION: Clone ID: ARATH-23APR03-CLUSTER13643_1			
US-10-739-930-227			
Query Match 53.7%; Score 1968.4; DB 20; Length 2679;			
Best Local Similarity 80.5%; Pred. No. 0;			
Matches 2657; Conservative 0; Mismatches 11; Indels 632; Gaps 5;			
Qy	368	GATTTAACTTATCTACTCTCAAAATTCATTAACCTTAGACGACCAACAGTCTC	427
Db	2	GATTTAACTTATCTACTCTCAAAATTCATTAACCTTAGACGACCAACAGTCTC	61
Qy	428	TTCAATATGTAAACAGAACAAAGTTTGTGTAGTACCCTAAAGACACTCCCATGAAG	487
Db	62	TTCAATATGTAAACAGAACAAAGTTTGTGTAGTACCCTAAAGACACTCCCATGAAG	121
Qy	488	CTGTGAGTCACGTGCGCAATTGCTCTCTCCCAATTCATTAATGCGATTACACCGGCGA	547
Db	122	CTGTGAGTCACGTGCGCAATTGCTCTCTCCCAATTCATTAATGCGATTACACCGGCGA	181
Qy	548	CGACAAAGCTCCGAGCTGACCAACCACTCTCAACAATATGCTCCGCGAGAAATGGG	607
Db	182	CGACAAAGCTCCGAGCTGACCAACCACTCTCAACAATATGCTCCGCGAGAAATGGG	241
Qy	608	CGACCGCTTCTCTCCGACTTCAATTTCAACCTCGAATCTCTCTCTCTCTCTGCGCA	667
Db	242	CGACCGCTTCTCTCTCCGACTTCAATTTCAACCTCGAATCTCTCTCTCTCTCTGCGCA	301
Qy	668	CGCGCACCAACGCGCACTCTGCTCTCTGCGCAACCATCTAATTGATCGTCCGAAAGCC	727
Db	302	CGCGCACCAACGCGCACTCTGCTCTCTGCGCAACCATCTAATTGATCGTCCGAAAGCC	361
Qy	728	ACGTCCCATCCCAATTAATTTCTAACCGAGTATTAAGAGCTCAACACATTTCTTAACG	787
Db	362	ACGTCCCATCCCAATTAATTTCTAACCGAGTATTAAGAGCTCAACACATTTCTTAACG	421
Qy	788	ATGGAATCAGAAAGACATTCGAGCTAGGGTTTCGAAACCGCGCGCAATTCGGTTTCAGG	847
Db	422	ATGGAATCAGAAAGACATTCGAGCTAGGGTTTCGAAACCGCGCGCAATTCGGTTTCAGG	481
Qy	848	ACGACGCTTTATCAAGCCGAGACAGATTTCTTCAAGCTGCTTTCGAAATCTCTGTCTAATC	907
Db	482	ACGACGCTTTATCAAGCCGAGACAGATTTCTTCAAGCTGCTTTCGAAATCTCTGTCTAATC	541
Qy	908	CTCGGCTCAGAAAGAGTAAATGAAGGCTTCTGATGATGAAGAGCTACAGTATCA	967
Db	542	CTCGGCTCAGAAAGAGTAAATGAAGGCTTCTGATGATGAAGAGCTACAGTATCA	601

QY	968	CTGATGTTCTCTGGGATAAGTAATTTTCGATTTCGGAAATATAAAGTTCTTCGTTTAA	1027
Db	602	CTGAATGTTCTCTGGGAT	618
QY	1028	TTTCATGAAATGGATAAAGGAAGAACTTTATCTAGTAAGAGTTCTCGGGGCTCTGT	1087
Db	619	-----MAGTTCCTGGTGCTCTGT	639
QY	1088	GTAATTGCAGAAGGTGGTGAGACTGAGATAGTCTTCGGGTGGTAGGCTTCGCTTAA	1147
Db	640	GTAATTGCAGAAGGTGGTGAGACTGAGATAGTCTTCGGGTGGTAGGCTTCGCTTAA	699
QY	1148	GAGAGGTTCCTTAAGTCCGTTTAAAGCAAGTGTGTTTAAATGAGCCCTTCGCTTTC	1207
Db	700	GAGAGGTTCCTTAAGTCCGTTTAAAGCAAGTGTGTTTAAATGAGCCCTTCGCTTTC	759
QY	1208	GATGCTCGAAGGAGTCTATGAGCATTTGATCCACCGTATTTTATACGTTATAGATT	1267
Db	760	GATGCTCGAAGGAGTCTATGAGCATTTGATCCACCGTATTTTATACGTTATAGATT	819
QY	1268	GTTGAGGAAGCTTTTGAAGCTTTTACAGGTAAGTTGACTTCGTTGGTAATTTGACGAGC	1327
Db	820	GTTGAGGAAGCTTTGAGCTTTTAC-----	844
QY	1328	TTGGCTTATATAGAACTTTCTGATTTGATTACTTTGATTGATCTTGCTGTAGAGGAA	1387
Db	845	-----AGAGAGAA	852
QY	1388	GGAGCAAGTACGCTTGACACCGGATTTACGTGACAAATTTGATGACCTTGGAGAGATC	1447
Db	853	GGAGCAAGTACGCTTGACACCGGATTTACGTGACAAATTTGATGACCTTGGAGAGATC	912
QY	1448	ACTCCGCGTTATGCTTGGAGCTACCTTACCGCTGGTGATGATTAACGCTCGAAA	1507
Db	913	ACTCCGCGTTATGCTTGGAGCTACCTTACCGCTGGTGATGATTAACGCTCGAAA	972
QY	1508	AGACTTAATGTTTAAAGCGGTGTGCGGAATATTTTGTGCTGTGTGAGAGGTGAGCA	1567
Db	973	AGACTTAATGTTTAAAGCGGTGTGCGGAATATTTTGTGCTGTGTGAGAGGTGAGCA	1032
QY	1568	TCAGCTCTGTTGGGGGTTTGAACCGGTGAGAGTTATGAATGAGCGCTTTTACGATG	1627
Db	1033	TCAGCTCTGTTGGGGGTTTGAACCGGTGAGAGTTATGAATGAGCGCTTTTACGATG	1092
QY	1628	ACAGCTGCTGACGAGGTATACAGTTAGATACCTTTTATATTTCTTACATGATATA	1687
Db	1093	ACAGCTGCTGAGC-----	1105
QY	1688	ACTTAGGTTTCTCATTTTATGATGTTGTGTGTGAGGTGATCTTTTGTAGCTACCC	1747
Db	1106	-----AGTTGATCTTTTGTAGCTACCC	1129
QY	1748	CAGCAATATTCACACAGATCATTTGAAGTTTAAAGTTGACATTCCTGCTGTGCTC	1807
Db	1130	CAGCAATATTCACACAGATCATTTGAAGTTTAAAGTTGACATTCCTGCTGTGCTC	1189
QY	1808	AAGCTTTATTTGGTAAAGACCAACACTTTTACAGAGTGTGATPAAGCAATTCACAGAAC	1867
Db	1130	AAGCTTTATTTGGTAAAGACCAACACTTTTACAGAGTGTGATPAAGCAATTCACAGAAC	1249
QY	1868	TTGACAGCGCTAAGGTAATGGCTATGAGATCTCTGCAGATTTGTATGATACACGGAATA	1927
Db	1250	TTGACAGCGCTAAGGTAATGGCTATGAGATCTCTGCAGATTTGTATGATACACGGAATA	1309
QY	1928	ATTGGGAATATGACTTCGCTCTAGAAAGGGACTCTGTGCACCTGTTATAGGCAAAAGTTG	1987
Db	1310	ATTGGGAATATGACTTCGCTCTAGAAAGGGACTCTGTGCACCTGTTATAGGCAAAAGTTG	1369
QY	1988	ATGAATGCCGATGTGTGTGGGCTTGAACGTGAGAGATTCACAATATAGGAATCCAGCTA	2047
Db	1370	ATGAATGCCGATGTGTGTGGGCTTGAACGTGAGAGATTCACAATATAGGAATCCAGCTA	1429
QY	2048	TTGTGAGATTTGTTTGGAGATTCATAATTCGTAGTACATATGATATCCCTGGAATAT	2107

Db	1430	TTGAGATTGTTTGGAGAAATTCGAATCGATATGACATGATGATCTCCCTGACAT	1489
Qy	2108	GCAAAATGTTGGAAACCTGTTGGCAGGGTGTCTTTCCTAGGTTCAAGACACCAAG	2167
Db	1490	GCAAAATGTTGGAAACCTGTTGGCAGGGTGTCTTTCCTAGGTTCAAGACACCAAG	1549
Qy	2168	ATAAAAAATTTAAACTGGGGGACATCATATGATCTCTATGCTTTTGAGTTACTTGGAA	2227
Db	1550	ATAAAAAATTTAAACTGGGGGACATCATATGATCTCTATGCTTTTGAGTTACTTGGAA	1609
Qy	2228	GAGTGGAGTGTGTCAGGGTCTCTCTTAGTGTGTCGCGCAACTATGGCAAGATTGGAG	2287
Db	1610	GAGTGGAGTGTGTCAGGGTCTCTCTTAGTGTGTCGCGCAACTATGGCAAGATTGGAG	1669
Qy	2288	CCGAGCATATGTAAAGCTAGTGTCTATGACGACATCGCAAGAAATTTTCTTCCCGCTATA	2347
Db	1670	CCGAGCATATGTAAAGCTAGTGTCTATGACGACATCGCAAGAAATTTTCTTCCCGCTATA	1729
Qy	2348	CAGATAGAACTCGGCTGAAACCCAGAGATGTGCAGAGACAGTGTTTAGTGTAGATCCTG	2407
Db	1730	CAGATAGAACTCGGCTGAAACCCAGAGATGTGCAGAGACAGTGTTTAGTGTAGATCCTG	1789
Qy	2408	TTTGCTTAACTATGTAGCCGCGATATGATGACCTGTGTCTTATTTGCAAGCTGTAGAC	2467
Db	1790	TTTGCTTAACTATGTAGCCGCGATATGATGACCTGTGTCTTATTTGCAAGCTGTAGAC	1849
Qy	2468	CCTCTGAAAACTTTGAAACCTATATGATATATGCAATTCAGCTGGGGTCTCAGAGATGACG	2527
Db	1850	CCTCTGAAAACTTTGAAACCTATATGATATATGCAATTCAGCTGGGGTCTCAGAGATGACG	1909
Qy	2528	TTTGATGAAACTACTGTTGAAATGTCCGTTGCTGATATGTTTAAAGAGCGAAGTGTGAAGA	2587
Db	1910	TTTGATGAAACTACTGTTGAAATGTCCGTTGCTGATATGTTTAAAGAGCGAAGTGTGAAGA	1969
Qy	2588	TTCCTTAGCTGCTGCTGTGGCAATTGGACTGATTTTCACTGTTCAAGCACAAGATTTTCTTA	2647
Db	1970	TTCCTTAGCTGCTGCTGTGGCAATTGGACTGATTTTCACTGTTCAAGCACAAGATTTTCTTA	2029
Qy	2648	AAAGCAGCTCATCTTTTCAAGCACAAGATATAGTTTCTTCTATGAAATCTGATCCCTTA	2707
Db	2030	AAAGCAGCTCATCTTTTCAAGCACAAGATATAGTTTCTTCTATGAAATCTGATCCCTTA	2089
Qy	2708	CCATAGGTATGATTTAAATGATGCAATTTTCAATATCTGCATTCGCAAAAATATGCTGT	2767
Db	2090	CCA-----	2092
Qy	2768	TTTGTGAGCTTAAAGAACTATGTTCCCACTTATATACATGTCCAAAAATGTTATCCAAGATT	2827
Db	2093	-----	2092
Qy	2828	ACAAGTTGCTGAGTAAATTTCACTAATATATGCTGTTGAATTTTGTATCAAACTGTAGA	2887
Db	2093	-----	2092
Qy	2888	CAGAAATGTAAATTTCACTCTCAATTTCTGTTAGAAATAACGTAGATYAGAGATTGC	2947
Db	2093	-----	2092
Qy	2948	CTTAGTGTGCTTTGTCCAATTTTCTTTCCTTGATTTTTCCTTTTTCGATTTTAGGTC	3007
Db	2093	-----TAGGTC	2100
Qy	3008	GTCAGACTGACGATTCAGAAAGCACTTCCAGATATGATGCTATGGACTGACAGAAATATA	3067
Db	2101	GTCAGACTGACGATTCAGAAAGCACTTCCAGATATGATGCTATGGACTGACAGAAATATA	2160
Qy	3068	GTAATCCAAAGTGGCAGAAGATTAATGCTCTGGCTTTTGGGCTGTATCCCGCATATGAATG	3127
Db	2161	GTAATCCAAAGTGGCAGAAGATTAATGCTCTGGCTTTTGGGCTGTATCCCGCATATGAATG	2220
Qy	3128	TTTACCAAGGTGAGGGAATTAATCTCAATTTCAATCAATTTGTGTGAAAACCTGTTGACAT	3187

Db 2221 TTACGAG----- 2227
 Qy 3188 GATTATAGTCTGGGCTTGTGTTGATTTATAGTTTGGATGGGCGAATGCT 3247
 Db 2228 -----AGGTTTTGGATGGGCGAATGCT 2249
 Qy 3248 GAAGATTGGACTGACAGAGAGCTGAAGCTGCGAGCTTGGGTTGATTATATAC 3307
 Db 2250 GAAGATTGGACTGACAGAGAGCTGAAGCTGCGAGCTTGGGTTGATTATAC 2309
 Qy 3308 ACTGTTGAACTATCTGTTGACAGTGTGACAGTCTGACAGATGAAACCCGCTGCTGT 3367
 Db 2310 ACTGTTGAACTATCTGTTGACAGTGTGACAGTCTGACAGATGAAACCCGCTGCTGT 2369
 Qy 3368 GGAAGCACTCTGAGAGAGTCTGTTGTTCTATCTGATTTGGTTATCCAGAAACATGC 3427
 Db 2370 GGAAGCACTCTGAGAGAGTCTGTTGTTCTATCTGATTTGGTTATCCAGAAACATGC 2429
 Qy 3428 TACTGATGTGAGACCTACACAAAGATAGAGATTTCTGTCGAAGTCAAGGTGAAA 3487
 Db 2430 TACTGATGTGAGACCTACACAAAGATAGAGATTTCTGTCGAAGTCAAGGTGAAA 2489
 Qy 3488 AATCACTGAAGGCTCTGTTCTTGATCATATATATCATATGAGCATGTGAGCTTG 3547
 Db 2490 AATCACTGAAGGCTCTGTTCTTGATCATATATATCATATGAGCATGTGAGCTTG 2549
 Qy 3548 CGAGATTCCTTTGTTCTGTTAATTTCTCTCTGATGTTAGTTTATTAATGAACAA 3607
 Db 2550 CGAGATTCCTTTGTTGTTGTAATTTCTCTCTGATGTTAGTTTATTAATGAACAA 2609
 Qy 3608 AAATTAAGCTTCTGGACACCCCTTTCTGATCTAACTATTAAGGCTCA 3667
 Db 2610 AAATTAAGCTTCTGGACACCAAAATCTTGAATCTAACTATTAAGGCTCA 2669

RESULT 4
 US-10-600-070-130
 ; Sequence 130. Application US/10600070
 ; Publication No. US20040139500A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Osteoryoung, Katherine W.
 ; APPLICANT: Vitsha, Stanislaw
 ; APPLICANT: Koksharova, Olga A.
 ; APPLICANT: Gao, Hongbo
 ; TITLE OF INVENTION: Placid Division and Related Genes and Proteins, and Methods of
 ; FILE OF INVENTION: Use
 ; FILE REFERENCE: MSU-08153
 ; CURRENT APPLICATION NUMBER: US/10/600,070
 ; CURRENT FILING DATE: 2003-06-20
 ; NUMBER OF SEQ ID NOS: 206
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 130
 ; LENGTH: 2637
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-10-600-070-130

Query Match 52.7%; Score 1931; DB 19; Length 2637;
 Best Local Similarity 80.4%; Pred. No. 0;
 Matches 2616; Conservative 0; Mismatches 5; Indels 632; Gaps 5;

Db 368 GATTTAATCTATCTACTCAAAATCAAAATTCCTAAACCTTAGACGACCAACAGCTC 427
 1 GATTTAATCTATCTACTCAAAATCAAAATTCCTAAACCTTAGACGACCAACAGCTC 60
 Qy 428 TTCAATATGTAACAGAAAGTTTGTGATGAGCTAAAGAGACCTCCATGGAAG 487
 Db 61 TTCAATATGTAACAGAAAGTTTGTGATGAGCTAAAGAGACCTCCATGGAAG 120
 Qy 488 CTCTGAGTCAGCTGCGCATTTGCTCTCCCATTCCAATTATGCGATTAACACCGCGA 547
 Db 121 CTCTGAGTCAGCTGCGCATTTGCTCTCCCATTCCAATTATGCGATTAACACCGCGA 180

Qy 548 CGACAAAGCTCCGACGTAGCCACAAACACTCTACAACTATCTGCTCCGACGAAATGGG 607
 Db 181 CGACAAAGCTCCGACGTAGCCACAAACACTCTACAACTATCTGCTCCGACGAAATGGG 240
 Qy 608 CCGACGCTTCTCTCCGACTTCAATTTCACTCCGATTCCTCTCTCTCTCTCTGCGCA 667
 Db 241 CCGACGCTTCTCTCCGACTTCAATTTCACTCCGATTCCTCTCTCTCTCTCTGCGCA 300
 Qy 668 CCGACGCTTCTCTCCGACTTCTGCTCTGCGCACTATGATGATGCTCCGAAAGCC 727
 Db 301 CCGACGCTTCTCTCCGACTTCTGCTCTGCGCACTATGATGATGCTCCGAAAGCC 360
 Qy 728 ACCTCCCATCCCATGATTTCTACAGGATATGAGAGCTCAACACATTTCTTAACG 787
 Db 361 ACCTCCCATCCCATGATTTCTACAGGATATGAGAGCTCAACACATTTCTTAACG 420
 Qy 788 ATGGAATCAGAAAGAGCTTGAAGCTTGAAGCGCGCAATTCGCTTCAAGC 847
 Db 421 ATGGAATCAGAAAGAGCTTGAAGCTTGAAGCGCGCAATTCGCTTCAAGC 480
 Qy 848 ACAGCGCTTTAATACCGCGAGACAGATTTCTGCAAGCTGCTTGGAAACTGCTATAC 907
 Db 481 ACAGCGCTTTAATACCGCGAGACAGATTTCTGCAAGCTGCTTGGAAACTGCTATAC 540
 Qy 908 CTGCGTCTAGAAAGAGATCAATGAAGTCTTCTTGAATGATGAAGAGCTACATCA 967
 Db 541 CTGCGTCTAGAAAGAGATCAATGAAGTCTTCTTGAATGATGAAGAGCTACATCA 600
 Qy 968 CTGATGTTCTTGGAGATGAATTTGATTTGATTTGGAATTAAGATTTCTGCTTTAA 1027
 Db 601 CTGATGTTCTTGGAGAT----- 617
 Qy 1028 TTTCATGAATTTGATTAAGAGAACTTTATCTAGTGAAGTTCTGCGGCTCTGCT 1087
 Db 618 -----AAGTCTCTGCTGCTCTGCT 638
 Qy 1088 GTATTGCAAGAGGTGTGAGACTGAGATGATTTCTGCGGTGTGAGGCTCTGCTTAAG 1147
 Db 639 GTATTGCAAGAGGTGTGAGACTGAGATGATTTCTGCGGTGTGAGGCTCTGCTTAAG 698
 Qy 1148 GAGAGGTGCTTAAAGTGTGTTAAGCAAGATGTGTTTAACTATGAGCGCTTCTCTC 1207
 Db 699 GAGAGGTGCTTAAAGTGTGTTAAGCAAGATGTGTTTAACTATGAGCGCTTCTCTC 758
 Qy 1208 GATGTCGAGAGGATGCTATGAGCTTGAATCCAGCTGATTTATCTGCTTAAAGATT 1267
 Db 759 GATGTCGAGAGGATGCTATGAGCTTGAATCCAGCTGATTTATCTGCTTAAAGATT 818
 Qy 1268 GTTGAAGAGCTTGAAGCTTTTACAGGTGATTGACTTGTCTTGTATTTAGCAGACG 1327
 Db 819 GTTGAAGAGCTTGAAGCTTTTAC----- 843
 Qy 1328 TTGGCTTTATTAAGAACTTCTTGAATTTGATATCTTGTATGAGTCTGTGTAGAGGAA 1387
 Db 844 -----ACGAGGAA 851
 Qy 1388 GAGCAAGATGCTTGTACCGGATTTACGTGACAAATTTGATGAGCTTTGAGAGAGATC 1447
 Db 852 GAGCAAGATGCTTGTACCGGATTTACGTGACAAATTTGATGAGCTTTGAGAGAGATC 911
 Qy 1448 ACTCCGCTTATGCTTGAAGCTTACTTGGCTTACCGCTTGTGTATGATTCGCTGCGAAA 1507
 Db 912 ACTCCGCTTATGCTTGAAGCTTACTTGGCTTACCGCTTGTGTATGATTCGCTGCGAAA 971
 Qy 1508 AGACTAAATGTTTAAACCGGTGTGCGGAATATTTGTGCTGTGTGAGAGAGGTGAGCA 1567
 Db 972 AGACTAAATGTTTAAACCGGTGTGCGGAATATTTGTGCTGTGTGAGAGAGGTGAGCA 1031
 Qy 1568 TCAAGCTTTGTGGGGGTTTGAACCGGTGAGAGATTTATGATGAGGCTTTTACGATG 1627
 Db 1032 TCAAGCTTTGTGGGGGTTTGAACCGGTGAGAGATTTATGATGAGGCTTTTACGATG 1091
 Qy 1628 ACAGCTCTGAGCAGTATACAGTTTATGATACCTTTTATTAATTTCTTATGATGATATA 1687

Db 1092 ACAGCTGCTAGC----- 1104
Qy 1688 ACTTAGGTTTCATTTTAATGATGTGNGTAGGGTTGATCTTTTGTAGCTACC 1747
Db 1105 -----AGTTTATCTTTTGTAGCTACC 1128
Qy 1748 CAAGCAATATTCAGCAGAGATCATTTGAAAGTTTACGAAGTTGCACTTGCTTGCTG 1807
Db 1129 CAAGCAATATTCAGCAGAGATCATTTGAAAGTTTACGAAGTTGCACTTGCTTGCTG 1188
Qy 1808 AAGCTTTATTTGGTAAGAGCCACACTTTTACAGATGCTGATTAAGCAATTCAGCAAC 1867
Db 1189 AAGCTTTATTTGGTAAGAGCCACACTTTTACAGATGCTGATTAAGCAATTCAGCAAC 1248
Qy 1868 TTGAGCAGGCTAAGGTAAATGGCTATGAGATTCCTGAGATGTTGTATGATACAGATA 1927
Db 1249 TTGAGCAGGCTAAGGTAAATGGCTATGAGATTCCTGAGATGTTGTATGATACAGATA 1308
Qy 1928 ATTGGAGATAGACTTGCTGCTAGAAAAGGGACTCTGTGCACTGCTTATAGCAAGTTG 1987
Db 1309 ATTGGAGATAGACTTGCTGCTAGAAAAGGGACTCTGTGCACTGCTTATAGCAAGTTG 1368
Qy 1988 ATGAATGCCGTATGTGGTGGCTTACAGATGAGATTCACAAATATAGAAATCCAGCTA 2047
Db 1369 ATGAATGCCGTATGTGGTGGCTTACAGATGAGATTCACAAATATAGAAATCCAGCTA 1428
Qy 2048 TTGAGGATTTGTTTGGAGAAATCAATGCTGATGCAATGATGATCTCCCTGGAATAT 2107
Db 1429 TTGAGGATTTGTTTGGAGAAATCAATGCTGATGCAATGATGATCTCCCTGGAATAT 1488
Qy 2108 GCAAAATTTTGAAACCTGTGTGGCAGGGGTGTCTTCTAGAGTTACAGACACCAAG 2167
Db 1489 GCAAAATTTTGAAACCTGTGTGGCAGGGGTGTCTTCTAGAGTTACAGACACCAAG 1548
Qy 2168 ATAAAAATTTTAACTCGGGGACTACTATGATGATCTTATGTTTGAATTA 2227
Db 1549 ATAAAAATTTTAACTCGGGGACTACTATGATGATCTTATGTTTGAATTA 1608
Qy 2228 GAGTGGAGGTATGATGAGGTTCTCTTATGCTGCTGCAACTATGAGCAAGATTGAG 2287
Db 1609 GAGTGGAGGTATGATGAGGTTCTCTTATGCTGCTGCAACTATGAGCAAGATTGAG 1668
Qy 2288 CCGAGCATGTGAAAGCTAGTGTATGAGGCACTGCAAGAAATTTTCTTCCCGCTATA 2347
Db 1669 CCGAGCATGTGAAAGCTAGTGTATGAGGCACTGCAAGAAATTTTCTTCCCGCTATA 1728
Qy 2348 CAGATGAAATCTGGGCTGAAACCCAGAGATGTGCAAGAGCACTGTTTATGTTAGTTCCTG 2407
Db 1729 CAGATGAAATCTGGGCTGAAACCCAGAGATGTGCAAGAGCACTGTTTATGTTAGTTCCTG 1788
Qy 2408 TTGTTAAATGATGAGGCTGATGAGGCTGCTGCTTATTTAGGAGAGCTGTAAGAC 2467
Db 1789 TTGTTAAATGATGAGGCTGATGAGGCTGCTGCTTATTTAGGAGAGCTGTAAGAC 1848
Qy 2468 CCTCTGAAAATCTTGAATCAATATGATTAATTCAGCTGGGGTCTCAGAGATGACG 2527
Db 1849 CCTCTGAAAATCTTGAATCAATATGATTAATTCAGCTGGGGTCTCAGAGATGACG 1908
Qy 2528 TTGATGAATCTACTGTTGAATGTCCGTGCTGATATGTTAAAGAGGCAAGTGTGAAGA 2587
Db 1909 TTGATGAATCTACTGTTGAATGTCCGTGCTGATATGTTAAAGAGGCAAGTGTGAAGA 1968
Qy 2588 TTCTAGTGTGCTGAGTGGCAATTTGCACTGATTCAGGCAAGATTTTCTTA 2647
Db 1969 TTCTAGTGTGCTGAGTGGCAATTTGCACTGATTCAGGCAAGATTTTCTTA 2028
Qy 2648 AAAGCAGCTCATCTTTCAACGCAAGATATGTTTCTTATAGAAATCTGATGCTGCTA 2707
Db 2029 AAAGCAGCTCATCTTTCAACGCAAGATATGTTTCTTATAGAAATCTGATGCTGCTA 2088
Qy 2708 CCAATGATATGATTAATATGATGCAATTTTCAATATCTGCAATGCTCAAAAATATGCTTGT 2767
|||

Db 2089 CCA----- 2091
Qy 2768 TTGTGAGCTAAGACATAGTTCCTTAATACATGTCCTCAAAAGTTGTACGAATTA 2827
Db 2092 ----- 2091
Qy 2828 ACAAGTTGCTGAGTAAATTTCACTAAATATATGCTGTAATTTTGTGATCAAACTGTAGA 2887
Db 2092 ----- 2091
Qy 2888 CAGAAATGTAATTTCACTCTCAACATTTCTGTTAGAAATAGTAGATTAAGATTC 2947
Db 2092 ----- 2091
Qy 2948 CTATGATGCTTTGTTCACACTTTTCTTCTGATTTTCTTTTCGATTTAGGCTCA 3007
Db 2092 -----TAGGCTCA 2099
Qy 3008 GTCAAGCTGACGATTCAGAGACATTTCCCAATGAGATCTTAGACTGCAAGATATA 3067
Db 2100 GTCAAGCTGACGATTCAGAGACATTTCCCAATGAGATCTTAGACTGCAAGATATA 2159
Qy 3068 GTATCCAGTGGCAAGAAATTAAGTCTGCTGGCTTTGGGCTGATCACCGATAGAAATG 3127
Db 2160 GTATCCAGTGGCAAGAAATTAAGTCTGCTGGCTTTGGGCTGATCACCGATAGAAATG 2219
Qy 3128 TTACCAAGGTGAGGGAATTAATCTCAATTCATCAATGTTGTGAACCTGTTGACAT 3187
Db 2220 TTACCAAG----- 2226
Qy 3188 GATTAATATGCTGCTGCTGCTTGTGATTCGTATTTATAGCTTTTGGATGGGGAATGCT 3247
Db 2227 -----AGTTTGTGATGGGGAATGCT 2248
Qy 3248 GAAGATTTGACCTGACAGACAGACTGAAACTGGCAGCTGGTGGTTATGATTAATAC 3307
Db 2249 GAAGATTTGACCTGACAGACAGACTGAAACTGGCAGCTGGTGGTTATGATTAATAC 2308
Qy 3308 ACTGTGAAATCTATCTGTTGACAGTGTGACAGTCTGACAGATGAAACCCGCTCTGCT 3367
Db 2309 ACTGTGAAATCTATCTGTTGACAGTGTGACAGTCTGACAGATGAAACCCGCTCTGCT 2368
Qy 3368 GGAAGCACTCTGAGAGAGTCTGCTGTATCTATGATTTGTTCTATCCAGAAAATATGC 3427
Db 2369 GGAAGCACTCTGAGAGAGTCTGCTGTATCTATGATTTGTTCTATCCAGAAAATATGC 2428
Qy 3428 TACTGATGTCAAACTTACACAAAGATGAGAAATTTCTGTTCCAAAGTCAAGGTGGA 3487
Db 2429 TACTGATGTCAAACTTACACAAAGATGAGAAATTTCTGTTCCAAAGTCAAGGTGGA 2488
Qy 3488 AATCATGAAAGGCTGTTCTGATCATATATATCATATGATGATGATGATGATGATGATG 3547
Db 2489 AATCATGAAAGGCTGTTCTGATCATATATATCATATGATGATGATGATGATGATGATG 2548
Qy 3548 CGAGATTTCTTTTGTCTGTAATTTCTCTCTAAGTTAGTTTAAATGAACACAAA 3607
Db 2549 CGAGATTTCTTTTGTGTAATTTCTCTCTAAGTTAGTTTAAATGAACACAAA 2608
Qy 3608 AAAATTAAGCTTC 3620
Db 2609 AAAATTAAGCTTC 2621

RESULT 5

US-10-600-070-1
; Sequence 1, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osefyoung, Katherine W.
; APPLICANT: Vilha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hongbo
; TITLE OF INVENTION: Placid Division and Related Genes and Proteins, and Methods of

TITLE OF INVENTION: Use
FILE REFERENCE: MSU-08153
CURRENT APPLICATION NUMBER: US/10/600.070
CURRENT FILING DATE: 2003-06-20
NUMBER OF SEQ ID NOS: 206
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1
LENGTH: 2406
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-10-600-070-1

Query Match 47.0% Score 1724; DB 19; Length 2406;
Best Local Similarity 79.2% Pred. No. 0;
Matches 2406; Conservative 0; Mismatches 0; Indels 632; Gaps 5;

QY 481 ATGGAAGCTCTGAGTCAAGTCGGCATTTGCTCTCCCATTTCAATTATGCGATTACCA 540
DB 1 ATGGAAGCTCTGAGTCAAGTCGGCATTTGCTCTCCCATTTCAATTATGCGATTACCA 60
QY 541 CCGGCCAGCAAAAGCTCCGACGTAGCCACAACCTCTACACTATCTGCTCGCCAGC 600
DB 61 CCGGCCAGCAAAAGCTCCGACGTAGCCACAACCTCTACACTATCTGCTCGCCAGC 120
QY 601 AATGGGCGGACGGCTCTCTCCGACTTGAATTTCACTCCGATTCCTCCTCCTCC 660
DB 121 AATGGGCGGACGGCTCTCTCCGACTTGAATTTCACTCCGATTCCTCCTCCTCC 180
QY 661 TTGCGCACCGCCACACACCGGCACTCTGCTCTCTGCGACCACTATTTGATCGTCC 720
DB 181 TTGCGCACCGCCACACACCGGCACTCTGCTCTCTGCGACCACTATTTGATCGTCC 240
QY 721 GAAGCGCAGTCCCATCCCATTTGATTTTCAAGATTTAGAGCTCAAAACATTTTC 780
DB 241 GAAGCGCAGTCCCATCCCATTTGATTTTCAAGATTTAGAGCTCAAAACATTTTC 300
QY 781 TTAAACGATGGAATTCAGAAAGCAATGAAGCTAGGGTTTCGAAACCGCGCAATTCGGT 840
DB 301 TTAAACGATGGAATTCAGAAAGCAATGAAGCTAGGGTTTCGAAACCGCGCAATTCGGT 360
QY 841 TTGAGCGACGACGCTTTAATCAAGCCGAGACAGATTTCTCAAGCTGCTCGAAATCTG 900
DB 361 TTGAGCGACGACGCTTTAATCAAGCCGAGACAGATTTCTCAAGCTGCTCGAAATCTG 420
QY 901 TCTAATCTCGGTCTAGAAAGAGATACATGAAGGCTCTTGTGATGAAGAGCTACA 960
DB 421 TCTAATCTCGGTCTAGAAAGAGATACATGAAGGCTCTTGTGATGAAGAGCTACA 480
QY 961 GTCACTCACTGATGTTCTTGGGATTAAGTAAATTCGATTCGGAATTAAGTTTCTTC 1020
DB 481 GTCACTCACTGATGTTCTTGGGAT----- 504
QY 1021 GTTTAATTTCAATGAATTTGATAAAGAGAACTTTATCTAGTGAAGTTCTTGAGGC 1080
DB 505 -----AAGTTCCTGAGGC 518
QY 1081 TCTCTGTGATTCAGAAAGGAGTGTGAGACTGAGATAGTTCTTGAGGTTGAGGCTCT 1140
DB 519 TCTCTGTGATTCAGAAAGGAGTGTGAGACTGAGATAGTTCTTGAGGTTGAGGCTCT 578
QY 1141 GCTTAAGGAAGGTTGCTAAGTGTGTTAAGCAAGATGCTTTTATGTTAAGGCGCTGC 1200
DB 579 GCTTAAGGAAGGTTGCTAAGTGTGTTAAGCAAGATGCTTTTATGTTAAGGCGCTGC 638
QY 1201 GTTCTCGATGTCTCAGAGGATGCTATGAGCACTGATGATTTTATTTATTCATGCTTA 1260
DB 639 GTTCTCGATGTCTCAGAGGATGCTATGAGCACTGATGATTTTATTTATTCATGCTTA 698
QY 1261 TGAAGTTGTGAGGAAGCTTTGAAGCTTTTACAGATGATTTGACTGCTTGTGTAATTTG 1320
DB 699 TGAAGTTGTGAGGAAGCTTTGAAGCTTTTAC----- 730
QY 1321 ACGAGCGTTGGCTTTATTAAGAACTTTCTGATTTGATTACTTTGTTATGAGCTTTGTGTA 1380

DB 731 -----A 731
QY 1381 GAGGAAGAGCAAGTACCTTTCACCCGATTTACGTCGCAAAATTGATNGAGACTTTGCA 1440
DB 732 GAGGAAGAGCAAGTACCTTTCACCCGATTTACGTCGCAAAATTGATNGAGACTTTGCA 791
QY 1441 AGAGATCACTCCGCGTATGCTTGGAGCTACTTGAGCTTCCGCTGATGATTAACGC 1500
DB 792 AGAGATCACTCCGCGTATGCTTGGAGCTACTTGAGCTTCCGCTGATGATTAACGC 851
QY 1501 TCGGAAAAGACTAAATGTTTAAAGCGGTGCGGAATATTTGTGTCTGTGGAGAGG 1560
DB 852 TCGGAAAAGACTAAATGTTTAAAGCGGTGCGGAATATTTGTGTCTGTGGAGAGG 911
QY 1561 TGGAGCATCACTCTTGTGGGGGTTTGAACCGGTGAGAACTTATGAATAGAGGCTTTT 1620
DB 912 TGGAGCATCACTCTTGTGGGGGTTTGAACCGGTGAGAACTTATGAATAGAGGCTTTT 971
QY 1621 ACGAATGACAGCTGAGGAGGATACAGTTTAGATACCTTTTAAATTTCTTTAGCA 1680
DB 972 ACGAATGACAGCTGAGG- 991
QY 1681 TGAATATTAATTTAGTTTCAATTTAATGATATGTTGTGATGATCTTTTGTGA 1740
DB 992 -----AGTTGATCTTTTGTGA 1008
QY 1741 GCTAACCCCAAGCAATATTCAGACAGAGTCAATTTGAATTAAGAAATTCGACTGCTCT 1800
DB 1009 GCTAACCCCAAGCAATATTCAGACAGAGTCAATTTGAATTAAGAAATTCGACTGCTCT 1068
QY 1801 GTGGCTCAAGCTTTTATGTTGAAGAGCAACCTTTTACAGATGCTGATTAAGCAATTC 1860
DB 1069 GTGGCTCAAGCTTTTATGTTGAAGAGCAACCTTTTACAGATGCTGATTAAGCAATTC 1128
QY 1861 CAGCAACTTGACAGGCTAAGTAAATGCTATGAGATTCCTGCAATGTTGTATGATACA 1920
DB 1129 CAGCAACTTGACAGGCTAAGTAAATGCTATGAGATTCCTGCAATGTTGTATGATACA 1188
QY 1921 CGGAATATATGGAGATTAAGCTTGGTCTTGAAGAAAGGGAATCTGTGCACTGCTATAGGC 1980
DB 1189 CGGAATATATGGAGATTAAGCTTGGTCTTGAAGAAAGGGAATCTGTGCACTGCTATAGGC 1248
QY 1981 AAGTTGATGAATGCGCTAATGTTGGGCTTGAAGTGAAGATTCACATATATAGGAT 2040
DB 1249 AAGTTGATGAATGCGCTAATGTTGGGCTTGAAGTGAAGATTCACATATATAGGAT 1308
QY 2041 CCAAGTATGAGAGTTGTTTGAAGAAATTCAAATGCTGATGACAAATGATCTCCCT 2100
DB 1309 CCAAGTATGAGAGTTGTTTGAAGAAATTCAAATGCTGATGACAAATGATCTCCCT 1368
QY 2101 GGAATATGCAAAATTTGTGAAAACCTGTTGAGAGGGGTTGCTTTCTAGGTTCAAGAC 2160
DB 1369 GGAATATGCAAAATTTGTGAAAACCTGTTGAGAGGGGTTGCTTTCTAGGTTCAAGAC 1428
QY 2161 ACCAAGATATAAATTTAAATCGGGGACACTAGATGATCTCTATGTTTGAAGTTAC 2220
DB 1429 ACCAAGATATAAATTTAAATCGGGGACACTAGATGATCTCTATGTTTGAAGTTAC 1488
QY 2221 TTGGAAGAGTGAAGTATGAGGTTCTCTTTAAGCTGCTGTCGCACTATGGAAGG 2280
DB 1489 TTGGAAGAGTGAAGTATGAGGTTCTCTTTAAGCTGCTGTCGCACTATGGAAGG 1548
QY 2281 ATTGAGCCGAGCATGTTGAAGGCTATGCTATGAGGCACTGCAAGAAAGTTTCTCTCC 2340
DB 1549 ATTGAGCCGAGCATGTTGAAGGCTATGCTATGAGGCACTGCAAGAAAGTTTCTCTCC 1608
QY 2341 CGCTATACATATGAATCTGGGCTGAACCAAGGATGGAAGAGAGAGAGTATAGTGA 2400
DB 1609 CGCTATACATATGAATCTGGGCTGAACCAAGGATGGAAGAGAGAGTATAGTGA 1668
QY 2401 GATCTGTGTTGTAACAATGATAGCCGTGATGAGCTGCTGTTATTTGCAAGACT 2460

Db 1669 GATCCTGTTGGTAAACAATGTAAGCCCTGATGCTGCTGTTTATTCAGAAAGCT 1728
 Qy 2461 GTAAGACCTCTGAAAACTTGGAACTAATGATTAATGCAATTCAGCTGGGGTCTCAGAG 2520
 Db 1729 GTAAGACCTCTGAAAACTTGGAACTAATGATTAATGCAATTCAGCTGGGGTCTCAGAG 1788
 Qy 2521 AGTAGCCTTATGATAAATCTACTGTTGAAATGTCCTGCTGATATGTTAAAGAGGCAAGT 2580
 Db 1789 AGTAGCCTTATGATAAATCTACTGTTGAAATGTCCTGCTGATATGTTAAAGAGGCAAGT 1848
 Qy 2581 GTGAAGATCCAGTGTGCTGGTGGGCAATGGAATGCACTGATTCAGCTCAGCCAGAGATAT 2640
 Db 1849 GTGAAGATCCAGTGTGCTGGTGGGCAATGGAATGCACTGATTCAGCTCAGCCAGAGATAT 1908
 Qy 2641 TTTCTTAAAGACGCTCATCTTTTCAACGCAAGGATATGTTTCTTATGAAATCGAT 2700
 Db 1909 TTTCTTAAAGACGCTCATCTTTTCAACGCAAGGATATGTTTCTTATGAAATCGAT 1968
 Qy 2701 GTGGCTAACCATGATGATTAATGATGCAATTTTCATATATCGCATTTGCTCAAAATA 2760
 Db 1969 GTGGCTAACCA----- 1978
 Qy 2761 TGCTGTGTTTGTAGTAAAGACATAGTCCACTAATACATGTCCTCAAAAGTTTACC 2820
 Db 1979 ----- 1978
 Qy 2821 AAGATTAAACAAGTTGCTGAGTAATTTCACTAATTAATGCTGTAATTTTGTGATCAAA 2880
 Db 1979 ----- 1978
 Qy 2881 CTGTAGACAGAAATGTAATTTCACTCTCAACATTTCTGTTAGAATAGAGATTAAG 2940
 Db 1979 ----- 1978
 Qy 2941 AGATTGCTTAGTGTGGCTTTGTCCAACTTTTCTTCTGATTTTCTTTTGCATTT 3000
 Db 1979 ----- 1979
 Qy 3001 AGGGTCAGTACAGAGCTGACGATTCAGAAAGCACTTCCAGAAATGAGTCTAGACCTGACAG 3060
 Db 1980 AGGGTCAGTACAGAGCTGACGATTCAGAAAGCACTTCCAGAAATGAGTCTAGACCTGACAG 2039
 Qy 3061 GAATATATGATTCAGAGTGGCAAGAAATTAAGTCTGAGCTTTTGGGCTGATCACCGCAT 3120
 Db 2040 GAATATATGATTCAGAGTGGCAAGAAATTAAGTCTGAGCTTTTGGGCTGATCACCGCAT 2099
 Qy 3121 AGAAATGTTTCCAGAGGTGAGGGAAATTAATCTAATCAATCAATGTTGTGAAAACTGT 3180
 Db 2100 AGAAATGTTTCCAG----- 2113
 Qy 3181 TGGACATGATTAATAGTGTGCTGCTGTTGATTCGTATTAATAGTGTGATGGGC 3240
 Db 2114 -----AGTGTGATGGGC 2128
 Qy 3241 GAATGCTGAAGATTTGGACTGACAGACAGCTGAACTGGCGAGCTTGGGTTGATATG 3300
 Db 2129 GAATGCTGAAGATTTGGACTGACAGACAGCTGAACTGGCGAGCTTGGGTTGATATG 2188
 Qy 3301 AATTATACATGTTGAAATCTATCTGTTGAAGTGTGAAGTCTGAGCAATGGAACCCGCTG 3360
 Db 2189 AATTATACATGTTGAAATCTATCTGTTGAAGTGTGAAGTCTGAGCAATGGAACCCGCTG 2248
 Qy 3361 CTCTGTGGAAGCACTCTGGAAGAGTCTGCTGTCTATCTGATTTGGTTCAATCCAGAAA 3420
 Db 2249 CTCTGTGGAAGCACTCTGGAAGAGTCTGCTGTCTATCTGATTTGGTTCAATCCAGAAA 2308
 Qy 3421 ACAATGCTACTGATGTCAGAACCTTACACAACAGATACGAAGTTTCTGATCCAAAGTCAG 3480
 Db 2309 ACAATGCTACTGATGTCAGAACCTTACACAACAGATACGAAGTTTCTGATCCAAAGTCAG 2368
 Qy 3481 GGTGGAATAATCACTGAAGGCTCTGTTCTTGTGATCATATA 3518
 Db 2369 GGTGGAATAATCACTGAAGGCTCTGTTCTTGTGATCATATA 2406

RESULT 6
 US-10-600-070-9
 ; Sequence 9, Application US/10600070
 ; Publication No. US20040139500A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Osteoryoung, Katherine W.
 ; APPLICANT: Vitha, Stanislaw
 ; APPLICANT: Koksharova, Olga A.
 ; APPLICANT: Gao, Hongbo
 ; TITLE OF INVENTION: Placid Division and Related Genes and Proteins, and Methods of
 ; TITLE OF INVENTION: Use
 ; FILE REFERENCE: MSU-08153
 ; CURRENT APPLICATION NUMBER: US/10/600,070
 ; CURRENT FILING DATE: 2003-06-20
 ; NUMBER OF SEQ ID NOS: 206
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 9
 ; LENGTH: 2406
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 US-10-600-070-9

Query Match 46.3%; Score 1720.8; DB 19; Length 2406;
 Best Local Similarity 79.1%; Pred. No. 0;
 Matches 2404; Conservative 0; Mismatches 2; Indels 632; Gaps 5;

Qy 481 ATGGAAGCTCTGATGACGTCGCGCATTTGCTCTCCCATTCGAATTAAGCGATTACCA 540
 Db 1 ATGGAAGCTCTGATGACGTCGCGCATTTGCTCTCCCATTCGAATTAAGCGATTACCA 60
 Qy 541 CCGGCGACGACAAAGCTCCGACGTAGCACAACACCTCTTACAACTATCTGCTCCGACG 600
 Db 61 CCGGCGACGACAAAGCTCCGACGTAGCACAACACCTCTTACAACTATCTGCTCCGACG 120
 Qy 601 AATTGGGCGACCGCTCTCTCTCCGACTTCAATTTACCTCCGATTCCTCTCTCTCC 660
 Db 121 AATTGGGCGACCGCTCTCTCTCCGACTTCAATTTACCTCCGATTCCTCTCTCTCC 180
 Qy 661 TTGGCACCGGCGACGACGACGACGCTGCTCTGCGACCACTATTAATGATCGGCC 720
 Db 181 TTGGCACCGGCGACGACGACGACGCTGCTCTGCGACCACTATTAATGATCGGCC 240
 Qy 721 GAAGCGACGTCGCCCATTCGCCATTTGATCTTACAGATTAAGAGCTCAACACATTTTC 780
 Db 241 GAAGCGACGTCGCCCATTCGCCATTTGATCTTACAGATTAAGAGCTCAACACATTTTC 300
 Qy 781 TTAAACGATGGAATCAAGAGACATTCGAAGCTTAGGCTTTCGAACCGCGCAATTCGGT 840
 Db 301 TTAAACGATGGAATCAAGAGACATTCGAAGCTTAGGCTTTCGAACCGCGCAATTCGGT 360
 Qy 841 TTACGAGGACGACGCTTAAATCAGCCGAGACAGATTCCTCAAGCTGCTGCGAACTCTG 900
 Db 361 TTACGAGGACGACGCTTAAATCAGCCGAGACAGATTCCTCAAGCTGCTGCGAACTCTG 420
 Qy 901 TCTAATCTCGGCTGTAAGAGAGTCAATGAAGTCTTCTTGATGATGAAGAGCTACA 960
 Db 421 TCTAATCTCGGCTGTAAGAGAGTCAATGAAGTCTTCTTGATGATGAAGAGCTACA 480
 Qy 961 GTCATCACTGATGTTCTTGGGATTAAGTAATTCGATTTCCGAATTAATTAAGTTCTTC 1020
 Db 481 GTCATCACTGATGTTCTTGGGAT----- 504
 Qy 1021 GTTTTAATTTCAATGAATGGAATTAAGAAAGAACTTTTATCTAATGTAAGGTTCTGGGGC 1080
 Db 505 -----AAGTTCTTGGGGC 518
 Qy 1081 TCTCTGATTTGCAAGAGAGTGTGAGACTGATAGTTCTTCTGGGTTGTGAGGCTCT 1140
 Db 519 TCTCTGATTTGCAAGAGAGTGTGAGACTGATAGTTCTTCTGGGTTGTGAGGCTCT 578
 Qy 1141 GCTTAAGAGAGGTTGCTTAAGTGTGTTTAAGCAAGATGTGTTTAAGTATAGGCGCTTGC 1200

Db 579 GCTTAAGAGAGGTGCTAAGTGGTTTAAGCAGATGTGTTTTAGTTAAGGCGCTGC 638
Qy 1201 GTTTCCTGATGTCGAGGGAGTCTATGCGCATTTGATCCACTGATTTTATTACTGGTTA 1260
Db 639 GTTTCCTGATGTCGAGGGAGTCTATGCGCATTTGATCCACTGATTTTATTACTGGTTA 698
Qy 1261 TGAAGTTGTTGAGGACCTTGAAGCTTTTACAGTAAGTTTGAAGCTTGTGGTAATTG 1320
Db 699 TGAAGTTGTTGAGGACCTTGAAGCTTTTAC----- 730
Qy 1321 ACAGAGCTGGCTTTATAGAACTTTCTTGATTTGATACCTTGTATGAGCTTGTGTA 1380
Db 731 -----A 731
Qy 1381 GAGAGAGAGAGCAATGATGCTTGCACCGGATTTACGTGCACAAATTGATGAGCTTTGGA 1440
Db 732 GAGAGAGAGAGCAATGATGCTTGCACCGGATTTACGTGCACAAATTGATGAGCTTTGGA 791
Qy 1441 AGAGATGATCCCGGTTATGCTTGAAGCTACCTGGCTTACCGCTGGTGAATGATTAAGC 1500
Db 792 AGAGATGATCCCGGTTATGCTTGAAGCTACCTGGCTTACCGCTGGTGAATGATTAAGC 851
Qy 1501 TGGCAAAAAGACTAATGCTTTAAGCGGTGTCGGAATATTTTGTGCTGTTGAGAGAG 1560
Db 852 TGGCAAAAAGACTAATGCTTTAAGCGGTGTCGGAATATTTTGTGCTGTTGAGAGAG 911
Qy 1561 TGGAGCATCAGCTTGTGTGGGGGTTTGACCGGTGAGAAAGTTATGAAATGAGCGGTTTT 1620
Db 912 TGGAGCATCAGCTTGTGTGGGGGTTTGACCGGTGAGAAAGTTATGAAATGAGCGGTTTT 971
Qy 1621 ACCAATGACGCTGCTGAGCAGGTATACAGTTAGATACCTTTTAAATTTCTTTAGCA 1680
Db 972 ATGATATGACGCTGCTAGC----- 991
Qy 1681 TGAATATGATTTAGTTTCTCATTTTATATGATGTGTGTGAGTGTGATCTTTTGTGA 1740
Db 992 -----AGGTTGATCTTTTGTGA 1008
Qy 1741 GCTAACCCCAAGCAATATTCAGACAGAGTCATTTGAAGTTTACGAAGTTGCACTTGTCTT 1800
Db 1009 GCTAACCCCAAGCAATATTCAGACAGAGTCATTTGAAGTTTACGAAGTTGCACTTGTCTT 1068
Qy 1801 GTGCTCAGCTTTATTTGTTGAAGAGCCACCTTTTACAGATGTCGATGAAACAAATTC 1860
Db 1069 GTGCTCAGCTTTATTTGTTGAAGAGCCACCTTTTACAGATGTCGATGAAACAAATTC 1128
Qy 1861 CAGCAACTTCAGCAGGCTTAAGTAATGCTATGAGATTCCTGCGATGTTGTATGATACA 1920
Db 1129 CAGCAACTTCAGCAGGCTTAAGTAATGCTATGAGATTCCTGCGATGTTGTATGATACA 1188
Qy 1921 CGGAATATATTTGAGAGATAGACTTCGGTCTAGAAAGGGGACTCTGTGCACTGCTTATAGGC 1980
Db 1189 CGGAATATATTTGAGAGATAGACTTCGGTCTAGAAAGGGGACTCTGTGCACTGCTTATAGGC 1248
Qy 1981 AAAGTTATGAAATGCGTATGCTGTGGGCTTAACAGTGAAGATTCACAAATATAGAAAT 2040
Db 1249 AAAGTTATGAAATGCGTATGCTGTGGGCTTAACAGTGAAGATTCACAAATATAGAAAT 1308
Qy 2041 CCAGCTATTTGAGAGTTGTTTGAAGAAATCAATGCTGATGCAATGATGATCTCCCT 2100
Db 1309 CCAGCTATTTGAGAGTTGTTTGAAGAAATCAATGCTGATGCAATGATGATCTCCCT 1368
Qy 2101 GGAATATGCAAAATTTGAGAAACTGTGTGCAAGGGGTTGTCTTTCTTGAAGTTGAGAGAC 2160
Db 1369 GGAATATGCAAAATTTGAGAAACTGTGTGCAAGGGGTTGTCTTTCTTGAAGTTGAGAGAC 1428
Qy 2161 ACCCAAAATATAAAAATTTAAATCTGGGGGACTATGATGATCTTATGCTTTTGAAGTAC 2220
Db 1429 ACCCAAAATATAAAAATTTAAATCTGGGGGACTATGATGATCTTATGCTTTTGAAGTAC 1488
Qy 2221 TTGGAAGAGGTGAGGTAGTTCAAGGTTCTCTTTAGCTGTGCTGCAACTATGCAAG 2280

Db 1489 TTGGAAGAGGTGAGGTAGTTCAAGGTTCTCTTTAGCTGTGCTGCAACTATGCAAG 1548
Qy 2281 ATTGAGCCGAGCATGAGAAAGCTAGTATGAGGCACTGCAAGAAAGTTTTCCTTCC 2340
Db 1549 ATTGAGCCGAGCATGAGAAAGCTAGTATGAGGCACTGCAAGAAAGTTTTCCTTCC 1608
Qy 2341 CGCTATACAGATAGAAACTGGCTGAAACCAAGATGTCAGAGACAGTGTTAGTGA 2400
Db 1609 CGCTATACAGATAGAAACTGGCTGAAACCAAGATGTCAGAGACAGTGTTAGTGA 1668
Qy 2401 GATCTGTTGGTAAACAATGATGAGCGGTGATGAGTGAAGCTGTGTTATTCAGAAAGCT 2460
Db 1669 GATCTGTTGGTAAACAATGATGAGCGGTGATGAGTGAAGCTGTGTTATTCAGAAAGCT 1728
Qy 2461 GTAAGACCTCTGAAACTTTGAAACTTAATGCAATTCAGAGCTGGGGGTCTCAGAG 2520
Db 1729 GTAAGACCTCTGAAACTTTGAAACTTAATGCAATTCAGAGCTGGGGGTCTCAGAG 1788
Qy 2521 AGTAGCCTGATGAAACTACGTGTTGAATGTCGCTGCTGATATGTTAAAGAGGCAAGT 2580
Db 1789 AGTAGCCTGATGAAACTACGTGTTGAATGTCGCTGCTGATATGTTAAAGAGGCAAGT 1848
Qy 2581 GTGAAGATCCTAGCTGCTGTGTGCAATTGCACTGATTCAGCTTCCAGCCAGAGATAT 2640
Db 1849 GTGAAGATCCTAGCTGCTGTGTGCAATTGCACTGATTCAGCTTCCAGCCAGAGATAT 1908
Qy 2641 TTTCTTAAAGAGCTCATCTTTTCAACGCAAGATATGCTTTCTTATGCAATCGAT 2700
Db 1909 TTTCTTAAAGAGCTCATCTTTTCAACGCAAGATATGCTTTCTTATGCAATCGAT 1968
Qy 2701 GTGCTACCAATGATGATTAATGATGCAATTTTCAATATCTGATTCGCTCAAAATA 2760
Db 1969 GTGCTACCA----- 1978
Qy 2761 TGTCTGTTTTGTGAGCTAAGAACATGATCCCACTTAATACATGCCAATAAGTTGTACC 2820
Db 1979 ----- 1978
Qy 2821 AAGATTAACAAGTTGCTGAGTAATTTCACTAATATATGCTGCTGAATTTTGTGATCAA 2880
Db 1979 ----- 1978
Qy 2881 CTGTAGACAGAAATGTAATTTCACTCTCAACATTTCTGTTAGATTAACGTAGAAATGAG 2940
Db 1979 ----- 1978
Qy 2941 AGATTCCTTAGTGTGCTTGTCCAACTTTTCTTCTTGATTTTCTTTTGCATTT 3000
Db 1979 -----T 1979
Qy 3001 AGGCTCAGTCAGAGCTGACGATTCAGAAAGCACTTCCAGAAATGATAGCTAGAGACTGAGA 3060
Db 1980 AGGCTCAGTCAGAGCTGACGATTCAGAAAGCACTTCCAGAAATGATAGCTAGAGACTGAGA 2039
Qy 3061 GAATATGATATCCAAAGTGGGCAAGATTAAGTCTCTGCTTTTGGGCTGATCACCAGAT 3120
Db 2040 GAATATGATATCCAAAGTGGGCAAGATTAAGTCTCTGCTTTTGGGCTGATCACCAGAT 2039
Qy 3121 AGAAATGTTACAGAGGTGAGGGAATTAATCTACAAATTCATCAATGTGTGAAAACTGT 3180
Db 2100 AGAAATGTTACAG----- 2113
Qy 3181 TGAAGATGATTAATGCTGTGCTTGTGATTTCTGTATTAATGATTTTGAATGGGC 3240
Db 2114 -----AGTTTTGAATGGGC 2128
Qy 3241 GAATGCTGAAGATTTGAGCTGACAGAGAGCTGAAACTGGCAGCTTGGGTGTTATG 3300
Db 2129 GAATGCTGAAGATTTGAGCTGACAGAGAGCTGAAACTGGCAGCTTGGGTGTTATG 2188
Qy 3301 ATTATACACTGTTGAAACTATCTGTGTAACAGTGTGACAGTCTCAGCAGATGAAACCCGTG 3360
Db 2189 ATTATACACTGTTGAAACTATCTGTGTAACAGTGTGACAGTCTCAGCAGATGAAACCCGTG 2248

QY 3361 CTCTGGTGAAGAACTCTGAGAGATCTGCTGTATCTATCTGATTTTGGTTCATCCAGAAA 3420
DB 2249 CTCTGGTGAAGAACTCTGAGAGATCTGCTGTATCTATCTGATTTTGGTTCATCCAGAAA 2308
QY 3421 ACAATGCTACTGATGTGACAACTACACAAAGATTAAGAAATTTTCTGTGCTCAAGTCAAG 3480
DB 2309 ACAATGCTACTGATGTGACAACTACACAAAGATTAAGAAATTTTCTGTGCTCAAGTCAAG 2368
QY 3481 GGTGAAAAATCACTGAAGGCTCTGTTCTTCTGATCATATA 3518
DB 2369 GGTGAAAAATCACTGAAGGCTCTGTTCTTCTGATCATATA 2406

RESULT 7
US-10-600-070-128
/ Sequence 128, Application US/10600070
/ Publication No. US20040139500A1
/ GENERAL INFORMATION:
/ APPLICANT: Oseeryoung, Katherine W.
/ APPLICANT: Vilha, Stanislav
/ APPLICANT: Koksharova, Olga A.
/ APPLICANT: Geo, Hongo
/ TITLE OF INVENTION: Plaeid Division and Related Genes and Proteins, and Methods of
/ FILE REFERENCE: MSU-08153
/ CURRENT APPLICATION NUMBER: US/10/600,070
/ CURRENT FILING DATE: 2003-06-20
/ NUMBER OF SEQ ID NOS: 206
/ SOFTWARE: Patencin version 3.2
/ SEQ ID NO 128
/ LENGTH: 2406
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-10-600-070-128

Query Match 46.8%; Score 1717.6; DB 19; Length 2406;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 2402; Conservative 0; Mismatches 4; Indels 632; Gaps 5;

QY 481 ATGGAAGCTCTGAGTCACTGCGGATGCTGCTCTCCCATTTCCAAATTAATGCGATTAACA 540
DB 1 ATGGAAGCTCTGAGTCACTGCGGATGCTGCTCTCCCATTTCCAAATTAATGCGATTAACA 60
QY 541 CCGGCGACGACAAAGCTCCGACGAGGACACAACTCTCAACTATCTGCTCGGACAG 600
DB 61 CCGGCGACGACAAAGCTCCGACGAGGACACAACTCTCAACTATCTGCTCGGACAG 120
QY 601 AAATGGCCGACGCTTCTCTCGACTTCATTTCACTTCCGATTCCTCTCTCTCTCC 660
DB 121 AAATGGCCGACGCTTCTCTCGACTTCATTTCACTTCCGATTCCTCTCTCTCTCC 180
QY 661 TTGGCCACCGCCACACACACCGGACCTGCTGCTCTGCGACCATTTATTTGATCGGCC 720
DB 181 TTGGCCACCGCCACACACACCGGACCTGCTGCTCTGCGACCATTTATTTGATCGGCC 240
QY 721 GAAGCGACGCTCCCATTTCCATTTTCTTACAGGATTTAGAGGCTCAAAACATTTTC 780
DB 241 GAAGCGACGCTCCCATTTCCATTTTCTTACAGGATTTAGAGGCTCAAAACATTTTC 300
QY 781 TTAAACGATGAATTCAGAAAGCATTTGAAAGTATTTGAAACCGCGCAATTCGCT 840
DB 301 TTAAACGATGAATTCAGAAAGCATTTGAAAGTATTTGAAACCGCGCAATTCGCT 360
QY 841 TTCAAGCAGCAGCTTTTAAATCAAGCCGAGACATTTCTTCAAGTGTCTTGGAAACTCTG 900
DB 361 TTCAAGCAGCAGCTTTTAAATCAAGCCGAGACATTTCTTCAAGTGTCTTGGAAACTCTG 420
QY 901 TCTAATCTCGGTCTAGAGAGATTAACAATGAAGTCTTCTTGAATGAAGAGCTACA 960
DB 421 TCTAATCTCGGTCTAGAGAGATTAACAATGAAGTCTTCTTGAATGAAGAGCTACA 480
QY 961 GTGATCACTGATGTTCTTGGGATTAAGTATTTGATTTGGAATTAATAAGTTCTTC 1020

DB 481 GTGATCACTGATGTTCTTGGGATTAAGTATTTGATTTGGAATTAATAAGTTCTTC 504
QY 1021 GTTTTAATTTCAAGAAATGGAATAAAGAAAGAACTTTTATCTAGTAAGGTTCTGGGGC 1080
DB 505 -----AAGTTCTGTGTC 518
QY 1081 TCTCTGTATTTGCAAGAAAGTGTGAGACTGAGATAGTTCTTGGGTTGGTGAAGGCTCT 1140
DB 519 TCTCTGTATTTGCAAGAAAGTGTGAGACTGAGATAGTTCTTGGGTTGGTGAAGGCTCT 578
QY 1141 GCTTAAAGAGAGTGTGCTTAAGTGTCTTTAAGCAAGATGTGTTTAACTTATGCGCTTGC 1200
DB 579 GCTTAAAGAGAGTGTGCTTAAGTGTCTTTAAGCAAGATGTGTTTAACTTATGCGCTTGC 638
QY 1201 GTTTCGATGCTCGAGGAGTGTATGAGCTTGAATCCACTGATTTTATCTAGGTTA 1260
DB 639 GTTTCGATGCTCGAGGAGTGTATGAGCTTGAATCCACTGATTTTATCTAGGTTA 698
QY 1261 TGAGTTTGTGAAGAGCTTTGAAGCTTTTACAGGATGATTTGACTTGTGTAATTTG 1320
DB 699 TGAGTTTGTGAAGAGCTTTGAAGCTTTTAC----- 730
QY 1321 ACAGAGCTTGGCTTTATTAAGAACTTTCTGATTTGATCTTGTATGAGTCTGTGCTA 1380
DB 731 -----A 731
QY 1381 GGAGAAAGAGCAAGTAGCTTGCACCGGATTTACGTGACAAATGATGAGACTTTGGA 1440
DB 732 GGAGAAAGAGCAAGTAGCTTGCACCGGATTTACGTGACAAATGATGAGACTTTGGA 791
QY 1441 AGAGATCACTCCGCTTATGCTTGAAGCTACTTGGCTTACCGCTTGTGATGATTAAGC 1500
DB 792 AGAGATCACTCCGCTTATGCTTGAAGCTACTTGGCTTACCGCTTGTGATGATTAAGC 851
QY 1501 TGGAAAAAGCTTAAATGCTTTAAGCGGTGTGCGGAATTTTGTGCTTGTGAGAGAG 1560
DB 852 TGGAAAAAGCTTAAATGCTTTAAGCGGTGTGCGGAATTTTGTGCTTGTGAGAGAG 911
QY 1561 TGGAGATCACTCTTGTGGGGTTTGAACCGGTGAGAAATTTATGAATGAGCGTTTTC 1620
DB 912 TGGAGATCACTCTTGTGGGGTTTGAACCGGTGAGAAATTTATGAATGAGCGTTTTC 971
QY 1621 ACGAATGACAGCTGCTGAGCAGGATTAACGTTTAACTTTTAAATTTCTTTAGCA 1680
DB 972 ACGAATGACAGCTGCTGAGC----- 991
QY 1681 TGAATTAATCTTAAAGTTTCTCAATTTAATGATGTTGTGATGATCTTTTGTGA 1740
DB 992 -----AGGTTGATCTTTTGTGA 1008
QY 1741 GGTACCCCAAGCAATTTTCCAGAGAGTCAATTTGAAGTTTACGAAGTTGACTTGTCTT 1800
DB 1009 GGTACCCCAAGCAATTTTCCAGAGAGTCAATTTGAAGTTTACGAAGTTGACTTGTCTT 1068
QY 1801 GTGGCTCAAGCTTTTATTTGTAAGAAACCAACCTTTTACAGGATGCTGTAAGCAATTC 1860
DB 1069 GTGGCTCAAGCTTTTATTTGTAAGAAACCAACCTTTTACAGGATGCTGTAAGCAATTC 1128
QY 1861 CAGCACTTCAAGAGCTTAAGTAAATGCTTAATGAGATTTCCGAGATGTTGATGATACA 1920
DB 1129 CAGCACTTCAAGAGCTTAAGTAAATGCTTAATGAGATTTCCGAGATGTTGATGATACA 1188
QY 1921 CCGAATTAATTTGAGAGTAACTTCCGCTTGAAGAAAGGAGACTGTGCACTGTTATAGGC 1980
DB 1189 CCGAATTAATTTGAGAGTAACTTCCGCTTGAAGAAAGGAGACTGTGCACTGTTATAGGC 1248
QY 1981 AAGTTGATGAATGCGGTATGTTGGCTTGAAGACAGTGAAGATTCACAATATAGGAAT 2040
DB 1249 AAGTTGATGAATGCGGTATGTTGGCTTGAAGACAGTGAAGATTCACAATATAGGAAT 1308
QY 2041 CCAAGCTTATGAGAGTTTGTGTTTGAAGAAATTCAAATCGTGAATGATGATCTCCCT 2100

Db 1309 CCAAGTATGTGAGTTGTTTGGAGAAATTCAAATCGATGACATGATGATCTCCCT 1368
 Qy 2101 GGAATATGCAAAATTTGTTGAAACCTGTTGGACGGGTTGCTTCTTCTAGTTCAGAGAC 2160
 Db 1369 GGAATATGCAAAATTTGTTGAAACCTGTTGGACGGGTTGCTTCTTCTAGTTCAGAGAC 1428
 Qy 2161 ACCAAGATATTAATACTCGGGGACCTACTATGATGATCCTATGTTGTTGAGTTAC 2220
 Db 1429 ACCAAGATATTAATACTCGGGGACCTACTATGATGATCCTATGTTGAGTTAC 1488
 Qy 2221 TTGAAAAGAGTGAAGTGTGAGGTTCTCTTTAGCTGCTGCTGCACTATGAGCAAG 2280
 Db 1489 TTGAAAAGAGTGAAGTGTGAGGTTCTCTTTAGCTGCTGCTGCACTATGAGCAAG 1548
 Qy 2281 ATTGAGCCCAAGATGGAAGCTGATGAGGCACTGCAAGAAAGTTTCTCTCC 2340
 Db 1549 ATTGAGCCCAAGATGGAAGCTGATGAGGCACTGCAAGAAAGTTTCTCTCC 1608
 Qy 2341 CGCTATACAGATGAAACTCGGCTGAACCAAGATGCAAGAGACAGTGTAGTGA 2400
 Db 1609 CGCTATACAGATGAAACTCGGCTGAACCAAGATGCAAGAGACAGTGTAGTGA 1668
 Qy 2401 GATCTGTTGGTAAACAATGATGAGCCGTGATGAGCTGCTGCTGCTTATTCAGAACT 2460
 Db 1669 GATCTGTTGGTAAACAATGATGAGCCGTGATGAGCTGCTGCTGCTTATTCAGAACT 1728
 Qy 2461 GTAAAGACCTCTGAAACCTTGAACCTAATGATTAATGCAATTCAGCTGGGCTTCAGAG 2520
 Db 1729 GTAAAGACCTCTGAAACCTTGAACCTAATGATTAATGCAATTCAGCTGGGCTTCAGAG 1788
 Qy 2521 AGTAGCGTTGATGAACTACTGTTGAATGCTGCTGCTGATGATGTTAAAGAGGCAAGT 2580
 Db 1789 AGTAGCGTTGATGAACTACTGTTGAATGCTGCTGCTGATGATGTTAAAGAGGCAAGT 1848
 Qy 2581 GTGAAGATCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2640
 Db 1849 GTGAAGATCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1908
 Qy 2641 TTTCTTAAAGAGCTCACTTTTCAACGCAAGATGTTTCTTATGAGATTCGAT 2700
 Db 1909 TTTCTTAAAGAGCTCACTTTTCAACGCAAGATGTTTCTTATGAGATTCGAT 1968
 Qy 2701 GTGCTACCACTAGTATGATTAATGATGCAATTTGATATCTGATGCTCAATA 2760
 Db 1969 GTGCTACCACTAGTATGATTAATGATGCAATTTGATATCTGATGCTCAATA 1978
 Qy 2761 TGCTGTTTTGTGAGCTAGAGACATAGTCCACTTAATACATGCTCCAAAGTTGATCC 2820
 Db 1979 TGCTGTTTTGTGAGCTAGAGACATAGTCCACTTAATACATGCTCCAAAGTTGATCC 1978
 Qy 2821 AAGATTAAACAAGTTGCTGAGTAAATTTCACTAATTATGCTGCTGAAATTTTTCATCAA 2880
 Db 1979 TGCTGTTTTGTGAGCTAGAGACATAGTCCACTTAATACATGCTCCAAAGTTGATCC 1978
 Qy 2881 CTGTAGACAGAAATGTAATTTCACTCAATTTCTGTTAGAAATGAGTATGAG 2940
 Db 1979 TGCTGTTTTGTGAGCTAGAGACATAGTCCACTTAATACATGCTCCAAAGTTGATCC 1978
 Qy 2941 AGATTGCTTAGTGTGCTTGTTCACATTTTCTTCTGATTTTCTTCTTCTGATTT 3000
 Db 1979 TGCTGTTTTGTGAGCTAGAGACATAGTCCACTTAATACATGCTCCAAAGTTGATCC 1979
 Qy 3001 AGGCTCAGTCAAGCTGACGATTCAGAAAGCACTTCCAGAAATGATGATGATGATGACGA 3060
 Db 1980 AGGCTCAGTCAAGCTGACGATTCAGAAAGCACTTCCAGAAATGATGATGATGATGACGA 2039
 Qy 3061 GAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3120
 Db 2040 GAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2099
 Qy 3121 AGAAATGTTACCAAGGTGAGGAATTAATCTCAATTCATCAATGTTGTGAAGTGT 3180
 Db 2100 AGAAATGTTACCAAGGTGAGGAATTAATCTCAATTCATCAATGTTGTGAAGTGT 2113

Qy 3181 TGGACATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
 Db 2114 -----AGTTTGGATGGC 2128
 Qy 3241 GAATGCTGAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
 Db 2129 GAATGCTGAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2188
 Qy 3301 ATTAATACATGTTGAACTATCTGTTGACAGTGTGACAGTGTGACAGTGTGACAGTGTGACAGTGTG 3360
 Db 2189 ATTAATACATGTTGAACTATCTGTTGACAGTGTGACAGTGTGACAGTGTGACAGTGTGACAGTGTG 2248
 Qy 3361 CTCTGTTGAAAGCACTCTGAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3420
 Db 2249 CTCTGTTGAAAGCACTCTGAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2308
 Qy 3421 ACAATGCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3480
 Db 2309 ACAATGCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2368
 Qy 3481 GGTGAAATCACTGAAAGCTGTTCTTCTGATCATTA 3518
 Db 2369 GGTGAAATCACTGAAAGCTGTTCTTCTGATCATTA 2406
 RESULT 8
 US-10-600-070-132/c
 ; Sequence 132, Application US/10600070
 ; Publication No. US20040139500A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Osteoryoung, Katherine W.
 ; APPLICANT: Viltha, Stanislaw
 ; APPLICANT: Koksharova, Olga A.
 ; TITLE OF INVENTION: Placitid Division and Related Genes and Proteins, and Methods of
 ; FILE REFERENCE: MSU-08153
 ; CURRENT APPLICATION NUMBER: US/10/600,070
 ; CURRENT FILING DATE: 2003-06-20
 ; NUMBER OF SEQ ID NOS: 206
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 132
 ; LENGTH: 561
 ; TYPE: DNA
 ; ORGANISM: Arabidopsis thaliana
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (127)..(127)
 ; OTHER INFORMATION: n is a, c, g, or t
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (520)..(520)
 ; OTHER INFORMATION: n is a, c, g, or t
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (541)..(541)
 ; OTHER INFORMATION: n is a, c, g, or t
 ; US-10-600-070-132
 Query Match 11.0%; Score 402.2; DB 19; Length 561;
 Best Local Similarity 82.2%; Pred. No. 4.6e-95;
 Matches 514; Conservative 0; Mismatches 20; Indels 91; Gaps 1;

Qy 2970 TTTCTTCTGATGATTTTCTTCTTCTGATTTAGGCTCAGTCAAGCTGACGATTCAGAG 3029
 Db 534 TTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 475
 Qy 3030 CACTTCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3089
 Db 474 CACTTCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 415
 Qy 3090 AGTCTGCTTTTGGCTGATCAAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3149


```
Db 414 AGCTCGGCTTTGGCCCTGATCACCGCAATGAAATGTTACCG----- 370
Qy 3150 TCTACAAATCAATCAATTTGTGTGAAAACCTGTGACATGATTAATGCTGGTCTGTT 3209
Db 369 ----- 370
Qy 3210 TGAATCTGTAATTAATAGTTTGGATGGCGAAATGCTGAAGATTGGACTGACAGACA 3269
Db 369 -----AGTTTGGATGGCGAAATGCTGAAGATTGGACTGACAGACA 326
Qy 3270 GCTGAACCTGCGACGCTGGGTTGGTTATGATTAATACATCTGTGAACATTAATCTGTTGAC 3329
Db 325 GCTGAACCTGCGACGCTGGGTTGGTTATGATTAATACATCTGTGAACATTAATCTGTTGAC 266
Qy 3330 AGTGTGACAGTCTCAGACAGATGAAACCGTCTCTGTGGAGAACATCTGAGAGAGTCT 3389
Db 265 AGTGTGACAGTCTCAGACAGATGAAACCGTCTCTGTGGAGAACATCTGAGAGAGTCT 206
Qy 3390 GCTTGTCTATCTGATTTGGTTCAATCAGAAAACAATGCTACTGATGTCAGAACTTACACA 3449
Db 205 GCTTGTCTATCTGATTTGGTTCAATCAGAAAACAATGCTACTGATGTCAGAACTTACACA 146
Qy 3450 ACAAGATACGAAGTTTCTGCTCAGTCAAGGCTGGAATAATCACTGAAGCTCTGTTCTT 3509
Db 145 ACAAGATACGAAGTTTCTGCTCAGTCAAGGCTGGAATAATCACTGAAGCTCTGTTCTT 86
Qy 3510 GCATCAATATATATCTCATATGATGATGCTGAGCTTGCAGATTCCTTTGTTCTGTA 3569
Db 85 GCATCAATATATATCTCATATGATGATGCTGAGCTTGCAGATTCCTTTGTTGTTA 26
Qy 3570 ATTCTCTCTTAAGTTAGTTTAT 3594
Db 25 ATTCTCTCTTAAGTTAGTTTAT 1
```

```
RESULT 9
US-10-600-070-184
; Sequence 184, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteoryoung, Katherine W.
; APPLICANT: Viltha, Stanislav
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Hongbo
; TITLE OF INVENTION: Placrid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 184
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Prunus persica
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (21)..(21)
; OTHER INFORMATION: n is a, c, g, or t
US-10-600-070-184
```

```
Query Match 8.1%; Score 295.6; DB 19; Length 631;
Best Local Similarity 71.7%; Pred. No. 6,8e-67;
Matches 407; Conservative 0; Mismatches 149; Indels 12; Gaps 1;
Qy 1708 ATGATATGTTGTGATGTTGATCTTTTGTAGTACATCCCAAGCAATATTCAGCAGAG 1767
Db 64 ATGATGTCAGCTGAGCGATTGATTTATTTAGTACATCCCAAGTATATTCAGCGAGAA 123
Qy 1768 TCATTTGAAGTTTACGAAGTTGCACTGCTCTTGTGCTCAAGTTTATTTATTTGAAG 1827
Db 124 AGCTTGAAGTTTATGGGGTGGCTCTTGGCTTGTCTCAAGCTTGTGTTGTTAAAAA 183
```

```
Qy 1828 CCAACCTTTTACAGATGCTGATTAAGCAATTCAGCACTTACAGCGCTTAAGTATG 1887
Db 184 CTTCAATCACTTAAGATGCTGAAAACCTATTCCAGAACTTACAGCGCTTAAGTAAACA 243
Qy 1888 GCTATGAGATTTCTCGGATGTTGTATGATACACGGAATTAATGGGATTAAGATTCGGT 1947
Db 244 GCTGTAGACATTTCTTTGCAACCTATATATACAAAGAAACAGTGAATGATGATTTGCT 303
Qy 1948 CTAGAAAAGGGAAGCTGTGCACTGCTTATAGCAAAAGTTGATGAATCCGTATGCTGTT 2007
Db 304 TTGGAAGGGAAGCTGTGCTTATCTTTCTTGAAGGGAAGCTTATGACAGTCTTGGTGTG 363
Qy 2008 GCGTTAGACAGTGAAGATTCACAAATATAGGAATCCAGCTATTGTGAGTTGTTGGAG 2067
Db 364 GCGCTAGACAGTATGATTCACCAATATAGAAATCCATCTGTGTAGACTTTGTCTGGAG 423
Qy 2068 AATTCAATATGATG-----ACAATGATGATCCCTGTGACATTAAGCAATG 2115
Db 424 AACTCAAAAGGATGACATGACAAATGACAAATGATGATCTTGTGACTTTCAGACTTA 483
Qy 2116 TTGAAAACCTGTTTGGCAGGGGTTGCTTTCTTAGGTTACAGACACCAAGATTAATAA 2175
Db 484 TTGGAAGCTGTGTTGATGAGAGTGTATTTCCAGGTTTATGAGACACCAAGATTAAG 543
Qy 2176 TTTAAACTCGGGGACTATATGATGATCTATGATGTTTGAAGTTGTTGAAGAGTGAAG 2235
Db 544 TTCAGACTGGGAAGCTATATGATGATCTTACAGTCTTGAATATTAAGAAAGCTGAT 603
Qy 2236 GTAGTTCAAGGTTCTCTTATGCTGCTG 2263
Db 604 GGCACTAATGTTTACCCCTTAGCTGCTG 631
```

```
RESULT 10
US-10-424-599-129007
; Sequence 129007, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 129007
; LENGTH: 1146
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87499C.1
US-10-424-599-129007
```

```
Query Match 7.2%; Score 262.4; DB 18; Length 1146;
Best Local Similarity 58.7%; Pred. No. 5,8e-58;
Matches 630; Conservative 0; Mismatches 266; Indels 178; Gaps 3;
Qy 587 TCTGCTCCGCGACAAATGCGGACCGCTTCTCTCCGACTTCAATTGACCTCCGATT 646
Db 236 TATCCGCAACACAAATATGGCGAGGCTCATTTGCGATTTCAATTCTCGGCGAGC 295
Qy 647 CTTCTCTCTCTCTCTTGGACCGCACACACCGGCACTCTGCTCTCTGCGACCAT 706
Db 296 CCGCGGCTTCCACTT-----CCACCTCCACTCTCAAGCCCTCTCTCTGCTCTCC 346
Qy 707 CTAATGATGTCCGGAAGCGCAGTCCCATCCCATATGATTTCTAACAGTATTAAGAG 766
Db 347 GCCTGATCTCTTGAAGCGTTAGCTGATCCCTCTTCAAGCTGTACCGCATTTCTGGCG 406
```



```
; GENERAL INFORMATION:
; APPLICANT: Fincher, Karen L.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51770)B
; CURRENT APPLICATION NUMBER: US/09/732,627A
; CURRENT FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 4930
; SEQ ID NO 2154
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(439)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3493-032-PI-M1-F8
; US-09-732-627A-2154

Query Match          4.8%; Score 175; DB 11; Length 439;
Best Local Similarity 65.0%; Pred. No. 3.7e-35;
Matches 278; Conservative 0; Mismatches 141; Indels 9; Gaps 1;

QY 499 GTCCGCAATTGGTCTCTCCCATTCATTTATGCGATTACCAACGCGAGCAAGCTC 558
   |||||
DB 12 GTTAGCATGGGGCTGTGACCTCGACACTTGTCTTATTGACCCGTGTAACCCCTCC 71
QY 559 CGACGTACCAACCAACCTCTACAACTATCTGTCCGCAAGATGGGCGAGCGCTCT 618
   |||||
DB 72 ANACTCCACACCGTTCAACCGTGTGTCTCCGCAAGTAATGGGCGAGCGCTCT 131
QY 619 CTCTCCGACTGATTTTCACTCCG-----ATTCTCTCTCTCTCTCTCTCTCTG 669
   |||||
DB 132 CTGCTGATCTTCATTTCTCTCCGCGCCGCAATATTCGTTCTCTCTCTCTCTCA 191
QY 670 GCCACCAACACCGCACTCTGTCTCTGCAACATTTATGATTCGCGCAAGCGCAC 729
   |||||
DB 192 ACCGCGACTCTTTTCTCCCTTACCTCCCTTCTCTGCGCCCTCTCTCTCGCAAG 251
QY 730 GTCCCATCCCATTTGATTTTACAGGATTTAGAGAGCTCAACACATTTCTTAAC 789
   |||||
DB 252 GTTTCATATCCCTTGAATTTCTACAGGATTTAGAGCGAGACTCATTTCTTAGT 311
QY 790 GGATTCAGAGAGATTTGAGGATTTGCGAAACCGCGCAATTCGTTTCAAGCAG 849
   |||||
DB 312 GGAATTTGAAGAGCTTATGAGCAAGGATTTGAAACCGCTCAATATGAGTTCA 371
QY 850 GACGCTTTAATCAGCGGAGAGAGATTTCTTCAAGCTTTGCGAAACTCTGTAT 909
   |||||
DB 372 GACACCATTAATGAGCGCAAGACATTTCTTAGCTGCTGTGAACCTTATACCT 431
QY 910 CGGTCTAG 917
   |||||
DB 432 GGCTCTAG 439

RESULT 15
US-10-600-070-175
; Sequence 175, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Oosteryoung, Katherine W.
; APPLICANT: Vitsha, Stanislaw
; APPLICANT: Koksharova, Olga A.
; TITLE OF INVENTION: Placid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: NSU-09153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 175
```

```
; LENGTH: 545
; TYPE: DNA
; ORGANISM: Gossypium arboreum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (528)..(528)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (536)..(536)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (540)..(540)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-600-070-175

Query Match          4.4%; Score 161; DB 19; Length 545;
Best Local Similarity 69.6%; Pred. No. 2.1e-31;
Matches 218; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 3224 ATAGCTTTGGATGGCGAATGCTGAAGATTTGACCTGACAGACAGCTGAAACTGCCCA 3283
   |||||
DB 82 AGAGTTCTGGATGATGCAATGTTGAAGACATGACAGATGTCAGCGCAATCGCTCA 141
QY 3284 GCTTGGTTGGTTATGATTATACACTGTTGAACATACTGTGACAGTGTGACAGTCTC 3343
   |||||
DB 142 GCTTGGTTGGTTATGATTATGATTATGATTATGATTATGATTATGATTATGATTATG 201
QY 3344 AGCAGATGAACCCGCTCTGTGTGAAGCAACTGTGAGAGTGTGCTGTATCTGA 3403
   |||||
DB 202 ACTAGATGGCAGGAGACTGTAGTGAAGTACTGTGAAGAAATCCACTGCTGACTGA 261
QY 3404 TTGCTTCATTCAGAAAACAATGCTACTGATGTCAAGAACTTACCAACAAGATACGA 3463
   |||||
DB 262 TGTTTCATTCATTCAGAAAACAATGCTACTGATGTCAAGAACTTACCAACAAGAT 321
QY 3464 TTGCTTCATTCAGAAAACAATGCTACTGATGTCAAGAACTTACCAACAAGATACGA 3523
   |||||
DB 322 GTTGTTCATTCAGAAAACAATGCTACTGATGTCAAGAACTTACCAACAAGATACGA 381
QY 3524 TCATATGTAGCAT 3536
   |||||
DB 382 ATGTATTAAGCAT 394
```

Search completed: June 10, 2005, 01:37:37
Job time : 2087 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 9, 2005, 15:25:22 ; Search time 575 Seconds
(without alignments)
10435.177 Million cell updates/sec

Title: US-10-600-070-3

Perfect score: 3667
Sequence: 1 tgcctgcattagaggaat.....ctataacataagcgctacaa 3667

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PC/US COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57.6	1.6	7218	1 US-08-232-463-14	Sequence 14, Appl
2	54	1.5	1141	4 US-09-806-708B-22	Sequence 22, Appl
3	50.2	1.4	57280	4 US-09-949-016-11796	Sequence 11796, A
4	50.2	1.4	57280	4 US-09-949-016-12843	Sequence 12843, A
5	50.2	1.4	57280	4 US-09-949-016-12844	Sequence 12844, A
6	50.2	1.4	57280	4 US-09-949-016-12846	Sequence 12846, A
7	50.2	1.4	57280	4 US-09-949-016-13542	Sequence 13542, A
8	50.2	1.4	57280	4 US-09-949-016-13543	Sequence 13543, A
9	50.2	1.4	57280	4 US-09-949-016-13544	Sequence 13544, A
10	50.2	1.4	57280	4 US-09-949-016-13545	Sequence 13545, A
11	50.2	1.4	57280	4 US-09-949-016-14633	Sequence 14633, A
12	50.2	1.4	57280	4 US-09-949-016-14634	Sequence 14634, A
13	50.2	1.4	57280	4 US-09-949-016-14635	Sequence 14635, A
14	50.2	1.4	57280	4 US-09-949-016-14636	Sequence 14636, A
15	50.2	1.4	57280	4 US-09-949-016-14637	Sequence 14637, A
16	50.2	1.4	57280	4 US-09-949-016-14638	Sequence 14638, A
17	50.2	1.4	57280	4 US-09-949-016-14639	Sequence 14639, A
18	50.2	1.4	57280	4 US-09-949-016-14640	Sequence 14640, A
19	49.4	1.3	913	1 US-08-217-327-3	Sequence 3, Appl
20	49.4	1.3	913	1 US-08-885-970A-3	Sequence 3, Appl
21	49.4	1.3	913	1 US-08-298-687A-3	Sequence 3, Appl
22	49.4	1.3	913	1 US-08-530-787-2	Sequence 2, Appl
23	49.4	1.3	913	1 US-08-298-829-3	Sequence 3, Appl
24	49.4	1.3	913	2 US-08-787-335-2	Sequence 25, Appl
25	49.4	1.3	1984	1 US-07-885-970A-25	Sequence 25, Appl
26	49.4	1.3	1985	1 US-08-298-687A-25	Sequence 25, Appl
27	49.4	1.3	1985	1 US-08-298-829-25	Sequence 25, Appl

28	49.2	1.3	324	4 US-09-547-693-234	Sequence 234, Appl
29	48.4	1.3	152132	4 US-09-949-016-13845	Sequence 13845, A
30	48.4	1.3	152145	4 US-09-949-016-12371	Sequence 12371, A
31	47.4	1.3	1141	4 US-09-806-708B-22	Sequence 22, Appl
32	46.8	1.3	114793	4 US-10-148-806-3	Sequence 3, Appl
33	45	1.2	49818	4 US-09-949-016-12458	Sequence 12458, A
34	45	1.2	49829	4 US-09-949-016-14011	Sequence 14011, A
35	44.2	1.2	4403765	3 US-09-103-840A-2	Sequence 2, Appl
36	44.2	1.2	4411529	3 US-09-103-840A-1	Sequence 1, Appl
37	43.8	1.2	832	4 US-09-621-976-2813	Sequence 2813, Ap
38	43.8	1.2	832	4 US-09-621-976-2813	Sequence 2813, Ap
39	42.8	1.2	1055	4 US-09-806-708B-23	Sequence 23, Appl
40	42.8	1.2	60990	4 US-09-949-016-14080	Sequence 14080, A
41	42.4	1.2	23849	4 US-09-949-016-13813	Sequence 13813, A
42	42.4	1.2	39154	4 US-09-949-016-12384	Sequence 12384, A
43	42.4	1.2	39154	4 US-09-949-016-12801	Sequence 12801, A
44	42.4	1.2	39443	4 US-09-949-016-14326	Sequence 14326, A
45	42.4	1.2	39443	4 US-09-949-016-14327	Sequence 14327, A

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpc-F18
US-08-232-463-14
Query Match 1.6%, Score 57.6, DB 1, Length 7218;

Qy	Db	Sequence	Score	Length	Indels	Gaps	0;
Qy	Db	510 TCTCTCCCATTCGAATTATGCGATTACACGCGCGACGACAAAGCTCCGATGACCA	1.5%	54	DB 4;	Length 1141;	Best Local Similarity 10.0%; Pred. No. 0.00016;
Qy	Db	1078 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	79;	Conservative 311;	Mismatches 389;	Indels 8;	Gaps 2;
Qy	Db	570 CAAACACTCTACACTATCTGCTCCGCGACGAAATGGGCGGACCGCTTCTCCGACTT	2393	TTAGTGTAGATCCGTGTGTGTAACAATGTAGCCGCTGATGTAGGCTGTGTATTATG	2452		
Qy	Db	1138 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	67	KYRRWYNNKSRWYKKWYKKBANNTSBRYYARRKKDKTYBYMTNNKKGKGTGRH	126		
Qy	Db	630 CAATTTCACCTCGATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2453	CAGAAAGCTGTAAAGCCTCTGAA-AACTTGGAACTAATGATATGATGCAATTGAGCTGG	2511		
Qy	Db	1198 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	127	YMRKRAMBDIVDHHYVYVAMNNATWTCMDKDKDKRTKRWKKNNAATGMDDDTKHMMNN	186		
Qy	Db	690 CGTCTCTGCGACCATCTATGATGATGCTCCGACGACGCTCCCATCCATGATTT	2512	GTCTCAGAGTAGACGCTTGATGAACTACTAGTGTGAATGTCCGTTGCTGATGTTAAG	2571		
Qy	Db	1258 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	187	NGCBTVMVRYKTRDMSBKRMNYGMBWYKQWNSYDVYVYVWMDMCKRYARVWRVTG	246		
Qy	Db	750 CTACACAGTATAGAGACTCAACACACTTTCTTACCAGATGAAATGAAAGACTTCA	2572	GAGGCAAGTGTGAAGATCTTAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2631		
Qy	Db	1318 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY					
Qy	Db	810 AGCTAGGTTTCGAAACCGCGCAATCGGTTTCAGGACGACGCTTATGACCGGAG					
Qy	Db	1378 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY					
Qy	Db	870 ACAGATTCCTTCAGCTGCTT 889					
Qy	Db	1438 CCAATTCCTTCTATCTCTT 1457					
Qy	Db	US-09-806-708B-22					
Qy	Db	Sequence 22, Application US/09806708B					
Qy	Db	Patent No. 6784342					
Qy	Db	GENERAL INFORMATION:					
Qy	Db	APPLICANT: The University of British Columbia					
Qy	Db	TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants					
Qy	Db	FILE REFERENCE: 4810-58741					
Qy	Db	CURRENT APPLICATION NUMBER: US/09/806,708B					
Qy	Db	CURRENT FILING DATE: 2001-04-03					
Qy	Db	PRIOR APPLICATION NUMBER: US 60/147,133					
Qy	Db	PRIOR FILING DATE: 1999-08-04					
Qy	Db	NUMBER OF SEQ ID NOS: 23					
Qy	Db	SOFTWARE: PatentIn version 3.0					
Qy	Db	SEQ ID NO 22					
Qy	Db	LENGTH: 1141					
Qy	Db	TYPE: DNA					
Qy	Db	ORGANISM: Artificial sequence					
Qy	Db	FEATURE:					
Qy	Db	NAME/KEY: promoter					
Qy	Db	LOCATION: (1)..(1141)					
Qy	Db	OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAEI promoters					
Qy	Db	US-09-806-708B-22					

Db	247	RMENYVAMETRAHRRRRYNNNGMTBAAYAYRRMTNNNNNNNAKAKCKAKYGYMIRABVNSTC	306
Qy	2632	CAGAACTATTTTCTTAAAAAGCAGCTCATCTTTTCAACGCAAGATATGCTTTCTTATG	2691
Db	307	TTWKSXTTKEVTS CWANNCKRAGDADKDHKMWKMSAAMGYNNNNNNNNNTYTKKARHBAEW	366
Qy	2692	GAATCGATGTCGCGACCATAGTGATGTATTAATGATGCAATTTTCATATATCTGATG	2751
Db	367	DMVMSAKKKKMAANAHAHSSKKMTWTYKRTTWNNNNNNTTMMKRMAMWYMKMDMBGTYN	426
Qy	2752	CTCAAAATATGCTTTGTTGTGAGCTAAGAACATAGTCCCACTTAATCATGTCCCAAA	2811
Db	427	NNNNGRYYGWTTKKDKOMTYYKKWANNCKRABMHKTCTHNTTMMKKKTYMNNCYWKS	486
Qy	2812	AGTTGTACCAAGATTAAACAGTTGCTGTAGTAATTTCACTAATATATGCT- - - - -GCTT	2864
Db	487	MTNGKSHBBAAYVTWYMMWMMRRYAHANNNDMDYMKACTWYKYBVCCKMNNYAAMYTK	546
Qy	2865	GAATTTTGTGATCAAACTGTAGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTG	2924
Db	547	SSWNTSYRYWKTNNSWRMSDTRSMKRAANNYAABHGYKMNTRMBWSHTWBHBRAG	606
Qy	2925	AATAACGAGGATTAGAGATTGCTTGTAGTGTGCTTTGTCCAACTTTTCTCTGATT	2984
Db	607	AAHYMBMTPAKCHCMATWYKAKKYAGAGSNNNNNNNNNNNNNNNNATTCARDVYAA	666
Qy	2985	TTTTTCTTTGATTTAGGTTAGGGTCAGTCAAGCTGACGATTCAGAGCACTCCACAGATG	3044
Db	667	RMVMAANAKWYXYKBAANNAAYTHANNMWGCNNNATDTRRTWKNNNNAGTWKNNNN	726
Qy	3045	ATGCTAGACGTCGACAGAAATATAGATTCAGTAAGTCGACAGAAATTAAGTCTTG	3104
Db	727	NAKMAAANKNAALAAVKKAKHWRANWAMGMAADAAABTTDKNNNGAYTKYTTNN	786
Qy	3105	GACCTGATCATCCGATAGAAATGTATACAGAGGTGAGGAGAAATAATCTCAATTCATCA	3164
Db	787	NTYRGVVTNTAARDGMANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	846
Qy	3165	ATTGTGT 3171	
Db	847	AMWTNKW 853	

RESULT 3
 US-09-949-016-11796/c
 ; Sequence 11796, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 APPLICANT: VENTER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CLO01307
 CURRENT APPLICATION NUMBER: US/09/949, 016
 PRIOR FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 11796
 LENGTH: 57280
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-11796

Query Match 1.4%; Score 50.2; DB 4; Length 57280;
 Best Local Similarity 53.9%; Pred. No. 0.031;
 Matches 103; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY	565	AGGCACAAACCTCTACAATCTATCTGTCGGCAGGAAATGGCCGACCGTCTTCTCTCC	624
Db	13101	AGGCACATTCACACACAACTCTGCTTCACACATCCACCTTCACTTATTAAGCTCC	13042
QY	625	GACTTCAATTCACTTCGATTCCTCTCTCTCTTTCGCAACGCGCAACCAACGCGC	684
Db	13041	ACCATCAATATTCCTTATCACTCCACCAACCGCTTTCACATCAACACACACACAC	12982
QY	685	ACTCTGCTCTCTGCGCACCATCTATTGATGTCGCCAAGCGCAAGTCCCATCCCAATT	744
Db	12981	ACAAACCAACCTCACTATCACTGACTTCACACATCAATTACCATCACTCCACCAACAG	12922
QY	745	GATTTCACCA	755
Db	12921	CATCATACCA	12911

```

RESULT 4
US-09-949-016-12843/C
Sequence 12843, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12843
LENGTH: 57280
TYPE: DNA
ORGANISM: Human
US-09-949-016-12843

```

[illegible]

```

RESULT 5
US-09-949-016-12844/C
; Sequence 12844, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949, 016

```

```

: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 12844
: LENGTH: 57280
: TYPE: DNA
: ORGANISM: Human
US-03-949-016-12844

```

Query Match	1.4%	Score 50.2	DB 4	Length 57280
Best Local Similarity	53.9%	Pred. No. 0.031		
Matches 103	Conservative 0	Mismatches 88	Indels 0	Gaps 0
QY	565	AGGCACAACACCTCTTCAACTATCTGCTCCGCGACGAAATGAGCGACCGTCTTCTTCC	624	
Db	13101	AGCACATATCACACCAACAACGTGCTTCACACACATCACACCTTCACCTTATGTAGCTCC	13042	
QY	625	GACCTCAATTTACCTCCGATTCCTCTCTCTCTCTCTCTCTCTGCGACGCGACCAACCGCC	684	
Db	13041	ACCATCAATATCCCTTATCAAGTCACACACACCGATTTCACATCACACACCAACACC	12982	
QY	685	ACTCTGCTCTCTCTGCGACCATCTATATGATGTGTCGCCGAACGCGACGCTGCCCATTCGCCATT	744	
Db	12981	ACAACACCAACCTCATCTATATGACTTCACACATCAATTAATTCATCACTCCACACAAC	12922	
QY	745	GATTTCTACCA	755	
Db	12921	CATCATCACGA	12911	

RESULT 6
 US-09-949-016-12846/c
 ; Sequence 12846, Application US/09949016
 ; Patent No. 6812339
 ; GENERAL INFORMATION:
 ; APPLICANT: VENTER, J. Craig et al.
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 ; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 ; FILE REFERENCE: C001307
 ; CURRENT APPLICATION NUMBER: US/09/949,016
 ; CURRENT FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/241,755
 ; PRIOR FILING DATE: 2000-10-20
 ; PRIOR APPLICATION NUMBER: 60/237,768
 ; PRIOR FILING DATE: 2000-10-03
 ; PRIOR APPLICATION NUMBER: 60/231,498
 ; PRIOR FILING DATE: 2000-09-08
 ; NUMBER OF SEQ ID NOS: 207012
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 12846
 ; LENGTH: 57280
 ; TYPE: DNA
 ; ORGANISM: Human
 ; US-09-949-016-12846

```
Query Match          1.4%; Score 50.2; DB 4; Length 57280;
Best Local Similarity 53.9%; Pred. No. 0.031;
Matches 109; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
```

OY 565 AGCCAAACACCCTTCACAATCAGTCTGCCGCAAGAATTGGCGCACGTTCTCTCC 624

Dδ 13101 AGAACAATCCACGCACAAATAAGCTCTCACACCATACCATTAATTAGACTCC 13042

OY 625 GACCTTGATTTCACCTCCGAATTCCTCTCTCTCTGTGCGCACGCGCACACGCGCC 684

Dδ 13041 ACCATCAATAATCCCCTTCACAGTCCACCAACACCGCTTCACCATACCAACGACGACC 12982

QY	685	ACTCTCGCTCTCTCTGCAACCATCTATTGATCGCCGAAGCCAGCGTCCCATCCCATTT	744
Db	12981	ACAACCCACACTCAGTCATCATGACTTCCACCATCAATTATCCATCACTCCACCAAC	12922
QY	745	GATTTCTACCA	755
Db	12921	CATCATCAAGA	12911

```

RESULT 7
US-09-949-016-13542/c
: Sequence 13542, Application US/09949016
: Patent No. 6812339
: GENERAL INFORMATION:
: APPLICANT: VENTER, J. Craig et al.
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
: FILE REFERENCE: CLO01307
: CURRENT APPLICATION NUMBER: US/09/949, 016
: CURRENT FILING DATE: 2000-04-14
: PRIOR APPLICATION NUMBER: 60/241,755
: PRIOR FILING DATE: 2000-10-20
: PRIOR APPLICATION NUMBER: 60/237,768
: PRIOR FILING DATE: 2000-10-03
: PRIOR APPLICATION NUMBER: 60/231,498
: PRIOR FILING DATE: 2000-09-08
: NUMBER OF SEQ ID NOS: 207012
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 13542
: LENGTH: 57280
: TYPE: DNA
: ORGANISM: Human
: US-09-949-016-13542

```

	Query Match	Similarity	1.44; Score 50.2; DB 4; Length 57280;
	Best Local	53.9%; Pred. No. 0.031;	
	Matches 103; Conservative	0; Mismatches 88; Indels 0; Gaps 0;	
QY	565 AGCCACAACCTTTCACACTATCTGCTCGCGACGAAATGGCGCAGCGTCTTCTTCC	624	
Db	13101 AGCACCATATCCACACACAACTGCTCCACACCATATCACCCTTTCACCATTTATGAGCTCC	13042	
QY	625 GACTTCAATTTTCACCTCCGATTCCTCTCTCTCTCCCTTGGCGACGCGCACCAACCGCC	684	
Db	13041 ACCATTCATATATCCCTTATCATCGTCCACACACGCTTTACATACACACACACACACC	12982	
QY	685 ACTCTCGTCTCTCTGCGCACCATCTATTTGATCGTCCGAAAGCGACGTCGCCATTTCCCAT	744	
Db	12981 ACAACGACGACCTGCTATTCATGACTTCCACCATCAACATTTACCATCACTCCACGACAC	12922	
QY	745 GATTTCACCA	755	
Db	12921 CATCATCACA	12911	

```

RESULT 8
US-09-949-016-13543/c
: Sequence 13543, Application US/09949016
: Patent No. 6812339
:
: GENERAL INFORMATION:
:
: APPLICANT: VENTER, J. Craig et al.
:
: TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
:
: FILE REFERENCE: CL001307
:
: CURRENT APPLICATION NUMBER: US/09/949,016
:
: PRIOR FILING DATE: 2000-04-14
:
: PRIOR APPLICATION NUMBER: 60/241,755
:
: PRIOR FILING DATE: 2000-10-20
:
: PRIOR APPLICATION NUMBER: 60/237,768
:
: PRIOR FILING DATE: 2000-10-03
:
: PRIOR APPLICATION NUMBER: 60/231,498
:
: PRIOR FILING DATE: 2000-09-08
:
: NUMBER OF SEQ ID NOS: 207012

```

```

: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 13543
: LENGTH: 57280
: TYPE: DNA
: ORGANISM: Human
US-03-949-016-13543

```

	Query Match	Best Local Similarity	Score	DB	Length	Mismatches	Conservative	Indels	Gaps
	1.4%	53.9%	50.2	4	57280	0	0	0	0
			Pred. No. 0.031			88			
QY	565	AGCCACAAACCTCTTACACTATATCTGCTCGCCAGCAAAATGGGCGACCCGTCCTTCTCTCC	624						
Db	13101	AGGACCATATCCACCAACAACTGGCTTCACACCATATCACCCTTACGATTTATGAGCTTC	13042						
QY	625	GACTTCATTTTCACCTCCGATTCCTCTCCTCTCTTGCGCACCGCACCAACCGCC	684						
Db	13041	ACCATCAATATCCCTTATACGTTCACACCAACCGCTTTTACCATTCACACACACACAC	12982						
QY	685	ACTCTGCTCTCTGCGACCATCTATTTGATGTCGCCGAAGCCACGTCCTCCATCCCATTT	744						
Db	12981	ACAACACACGACTCATATCATGACTTCCACATCAATATACCATCATCCTCCACCAAC	12922						
QY	745	GATTTCTACCA	755						
Db	12921	CATCATCACA	12911						

RESULT 9
 US-09-949-016-13544/C
 Sequence 13544: Application US/09949016
 Patent No. 6812339
 GENERAL INFORMATION:
 APPLICANT: VENTNER, J. Craig et al.
 TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
 TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CL001307
 CURRENT APPLICATION NUMBER: US/09/949,016
 CURRENT FILING DATE: 2000-04-14
 PRIOR APPLICATION NUMBER: 60/241,755
 PRIOR FILING DATE: 2000-10-20
 PRIOR APPLICATION NUMBER: 60/237,768
 PRIOR FILING DATE: 2000-10-03
 PRIOR APPLICATION NUMBER: 60/231,498
 PRIOR FILING DATE: 2000-09-08
 NUMBER OF SEQ ID NOS: 207012
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 13544
 LENGTH: 57280
 TYPE: DNA
 ORGANISM: Human
 US-09-949-016-13544

	Query Match	Best Local Match	Similarity	Score	DB	Length	57280;
Matches	103;	Conservative	53.9%;	Pred. No. 0.031;	Mismatches	88;	Indels 0; Gaps 0;
QY	565	AGCCACAACCTCTTCAACTATCTGCTCCGACGAAATGGCGACCGCTCTTCTTCC					624
Db	13101	AGCACATTCACGACCAACCACTGCTCCACACCATCATCACACCTTCACGCTTATGAGCTCC					13042
QY	625	GACTTCAATTTGACCTCCGATTCCTCTCTCTCTCTGTCGACGCGGACGACCAACCGCC					684
Db	13041	ACCATCAATATCCCTATTCAGCTCACACACACGCTTTTCAACATACGACCAACACACACC					12988
QY	685	ACTCTGCTCTCTGCGACCATTCATTGATTCGTCCGAAGCGCACGCTCCCATTCGCCATT					744
Db	12981	ACAAACACACCTACATATATGACTTCCACGACATCAACATTACATCACTCCACGACACAC					12922
QY	745	GATTTCTACCA					755
Db	12921	CATCATCACGA					12911

RESULT 10

US-09-949-016-13545/C
; Sequence 13545, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13545
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13545

Query Match 1.4%; Score 50.2; DB 4; Length 57280;
Best Local Similarity 53.9%; Pred. No. 0.031;
Matches 103; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 565 AGCCACAACCTTACCACTATGCTCCGCGCAAGATGGCGGACCGCTTCTCTCC 624
DB 13101 AGCACCACTACACCAACCACTGCTCCACCACTACCACTTACCATTAAGCTCC 13042
QY 625 GACTTCAATTTGACCTCGATTCCTCTCTCTCTTGGCCAGCCAGCAGCAGCC 684
DB 13041 ACATCAATATCCCTATCAGCTCCACCAACCGCTTTCACATCAGCAGCAGC 12982
QY 685 ACTCTGCTCTCTGCAACCATTAATGATGTGCGGAGCCAGTCCCATCCCAT 744
DB 12981 ACAACCACTCACTATCATGACTTCCACCATCAATTAACATCTCCACCAAC 12922
QY 745 GATTTTACCA 755
DB 12921 CATCATCACGA 12911

RESULT 11

US-09-949-016-14633/C
; Sequence 14633, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14633
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14633

Query Match 1.4%; Score 50.2; DB 4; Length 57280;

Best Local Similarity 53.9%; Pred. No. 0.031;
Matches 103; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 565 AGCCACAACCTTACCACTATGCTCCGCGCAAGATGGCGGACCGCTTCTCTCC 624
DB 13101 AGCACCACTACACCAACCACTGCTCCACCACTACCACTTACCATTAAGCTCC 13042
QY 625 GACTTCAATTTGACCTCGATTCCTCTCTCTCTTGGCCAGCCAGCAGCAGCC 684
DB 13041 ACATCAATATCCCTATCAGCTCCACCAACCGCTTTCACATCAGCAGCAGC 12982
QY 685 ACTCTGCTCTCTGCAACCATTAATGATGTGCGGAGCCAGTCCCATCCCAT 744
DB 12981 ACAACCACTCACTATCATGACTTCCACCATCAATTAACATCTCCACCAAC 12922
QY 745 GATTTTACCA 755
DB 12921 CATCATCACGA 12911

RESULT 12

US-09-949-016-14634/C
; Sequence 14634, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14634
; LENGTH: 57280
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14634

Query Match 1.4%; Score 50.2; DB 4; Length 57280;
Best Local Similarity 53.9%; Pred. No. 0.031;
Matches 103; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 565 AGCCACAACCTTACCACTATGCTCCGCGCAAGATGGCGGACCGCTTCTCTCC 624
DB 13101 AGCACCACTACACCAACCACTGCTCCACCACTACCACTTACCATTAAGCTCC 13042
QY 625 GACTTCAATTTGACCTCGATTCCTCTCTCTCTTGGCCAGCCAGCAGCAGCC 684
DB 13041 ACATCAATATCCCTATCAGCTCCACCAACCGCTTTCACATCAGCAGCAGC 12982
QY 685 ACTCTGCTCTCTGCAACCATTAATGATGTGCGGAGCCAGTCCCATCCCAT 744
DB 12981 ACAACCACTCACTATCATGACTTCCACCATCAATTAACATCTCCACCAAC 12922
QY 745 GATTTTACCA 755
DB 12921 CATCATCACGA 12911

RESULT 13

US-09-949-016-14635/C
; Sequence 14635, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

```

: TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
:
: FILE REFERENCE: CLO01107
:
: CURRENT APPLICATION NUMBER: US/09/949,016
:
: CURRENT FILING DATE: 2000-04-14
:
: PRIOR APPLICATION NUMBER: 60/241,755
:
: PRIOR FILING DATE: 2000-10-20
:
: PRIOR APPLICATION NUMBER: 60/237,768
:
: PRIOR FILING DATE: 2000-10-03
:
: PRIOR APPLICATION NUMBER: 60/231,498
:
: PRIOR FILING DATE: 2000-09-08
:
: NUMBER OF SEQ ID NOS: 207012
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 14635
:
: LENGTH: 57280
:
: TYPE: DNA
:
: ORGANISM: Human
:
: US-09-949-016-14635

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 8, 2005, 18:54:19 ; Search time 1797 Seconds
(without alignments)
12079.962 Million cell updates/sec

Title: US-10-600-070-3

Perfect score: 3667

Sequence: 1 tgcctgcataaaggagac.....ctataacataaaggctacaa 3667

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3667	100.0	3667	ADJ38130	Adj38130 Arabidops
2	3663.8	99.9	3667	ADJ38136	Adj38136 Arabidops
3	1931	52.7	2637	ADJ38210	Adj38210 Plastid d
4	1724	47.0	2406	ADJ38129	Adj38129 Arabidops
5	1720.8	46.9	2406	ADJ38135	Adj38135 Arabidops
6	1717.6	46.8	2406	ADJ38208	Adj38208 Plastid d
7	402.2	11.0	561	ADJ38212	Adj38212 Plastid d
8	295.6	8.1	631	ADJ38264	Adj38264 Plastid d
9	208.8	5.7	660	ADJ38215	Adj38215 Plastid d
10	194	5.3	2283	ADJ38206	Adj38206 Plastid d
11	178.6	4.9	607	ADJ38781	Adj38781 Cotton pr
12	175	4.8	439	ADJ38143	Adj38143 Cotton ex
13	161	4.4	545	ADJ38255	Adj38255 Plastid d
14	159.4	4.3	552	ADJ38855	Adj38855 Cotton pr
15	158	4.3	608	ADJ38217	Adj38217 Plastid d
16	145.6	4.0	527	ADJ38214	Adj38214 Plastid d
17	136.2	3.7	642	ADJ38265	Adj38265 Plastid d
18	134.6	3.7	657	ADJ38266	Adj38266 Plastid d
19	132.2	3.6	563	ADJ38258	Adj38258 Plastid d
20	130.2	3.6	446	ADJ38231	Adj38231 Plastid d

21	127.6	3.5	479	12	ADJ38230	Adj38230 Plastid d
22	125	3.4	537	12	ADJ38223	Adj38223 Plastid d
23	122.6	3.3	307	12	ADJ38218	Adj38218 Plastid d
24	118.6	3.2	871	12	ADJ38232	Adj38232 Plastid d
25	93.6	2.6	491	12	ADJ38254	Adj38254 Plastid d
26	92.6	2.5	187	12	ADJ38216	Adj38216 Plastid d
27	92	2.5	418	12	ADJ38224	Adj38224 Plastid d
28	91.4	2.5	460	12	ADJ38256	Adj38256 Plastid d
29	90.4	2.5	620	12	ADJ38227	Adj38227 Plastid d
30	89.4	2.4	604	12	ADJ38237	Adj38237 Plastid d
31	89.4	2.4	653	12	ADJ38228	Adj38228 Plastid d
32	87.2	2.4	480	12	ADJ38225	Adj38225 Plastid d
33	86.8	2.4	622	12	ADJ38226	Adj38226 Plastid d
34	85.2	2.3	300	12	ADJ38260	Adj38260 Plastid d
35	81.8	2.2	360	12	ADJ38259	Adj38259 Plastid d
36	80.8	2.2	606	12	ADJ38257	Adj38257 Plastid d
37	78	2.1	416	12	ADJ38219	Adj38219 Plastid d
38	77.2	2.1	549	12	ADJ38261	Adj38261 Plastid d
39	75.8	2.1	540	12	ADJ38233	Adj38233 Plastid d
40	66.6	1.8	336	12	ADJ38222	Adj38222 Plastid d
41	65.6	1.8	309	12	ADJ38221	Adj38221 Plastid d
42	60.2	1.6	295	12	ADJ38213	Adj38213 Plastid d
43	59.4	1.6	400	12	ADJ38263	Adj38263 Plastid d
44	54	1.5	535	12	ADJ38229	Adj38229 Plastid d
45	51.6	1.4	2000	8	ADA71938	Ada71938 Rice Gene

ALIGNMENTS

RESULT 1
ADJ38130
ID ADJ38130 standard; DNA; 3667 BP.
XX
AC ADJ38130;
XX
DT 06-MAY-2004 (first entry)
XX
DE Arabidopsis thaliana Arabc6-1 genomic DNA SegID2.
XX
KW prokaryotic type; plastid division; Ftn2; ARc6; ARc5; Fzo; plant cell;
XX
KW agronomic; horticultural; crop plant; ornamental plant; woody plant;
XX
KW herbicide target; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO2004001003-A2.
XX
PD 31-DEC-2003.
XX
PF 20-JUN-2003; 2003WO-US019536.
XX
PR 20-JUN-2002; 2002US-0390140P.
XX
PR 09-AUG-2002; 2002US-0402242P.
XX
PR 20-JUN-2003; 2003US-00600070.
XX
PA (UNMS) UNIV MICHIGAN STATE.
XX
PI Osteryoung KM, Vitla S, Koksharova OA, Gao H;
XX
DR WPI: 2004-082486/08.
XX
DR P-PSDB; ADJ38202.
XX
PT New isolated Ftn2, ARc5 and/or Fzo-like nucleic acid sequences, useful
XX
PT for further characterizing plastid division in plant cells, and in
XX
PT varying agronomic and horticultural characteristics of economically
XX
PT important plants.
XX
PS Claim 1; SEQ ID NO 2; 287pp; English.
XX
CC This invention relates to novel prokaryotic type or plastid division and
XX
CC related genes and proteins. In particular, the invention relates to novel
XX
CC Ftn2 (ARc6), ARc5 and Fzo-like genes and polypeptides. The methods and

CC compositions of the present invention are useful for further
CC characterizing plant/division in plant cells, in order to vary
CC agronomic and horticultural characteristics of economically important
CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is that of a DNA sequence
CC which is related to the invention.

XX Sequence 3667 BP; 983 A; 670 C; 842 G; 1172 T; 0 U; 0 Other;

Query Match 100.0%; Score 3667; DB 12; Length 3667;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3667; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 TGTTCGATTAAGAGAAATCAATTAATACAAATTTGCTTGAATTTCAACAAATTTTG 60
DB 1 TGTTCGATTAAGAGAAATCAATTAATACAAATTTGCTTGAATTTCAACAAATTTTG 60
QY 61 CTGGGCTAATAGATTAATTCATGCTCTGTTGCTTTTACATTAATGATTAATGTTGG 120
DB 61 CTGGGCTAATAGATTAATTCATGCTCTGTTGCTTTTACATTAATGATTAATGTTGG 120
QY 121 AATTTACACATTTCAATGATGATTTAAGAAAGAGAGAAATGATGGGTTTGTGGG 180
DB 121 AATTTACACATTTCAATGATGATTTAAGAAAGAGAGAAATGATGGGTTTGTGGG 180
QY 181 TTTAACTTTAAAGTCAAGAAATTAAGTATGTTTACTGTTGCTCTAATATGTGAA 240
DB 181 TTTAACTTTAAAGTCAAGAAATTAAGTATGTTTACTGTTGCTCTAATATGTGAA 240
QY 241 AATGAAGCAATCCAAACGTTCTTAGTGAATAGATTAATTAAGAGATTAATCAT 300
DB 241 AATGAAGCAATCCAAACGTTCTTAGTGAATAGATTAATTAAGAGATTAATCAT 300
QY 301 AAAGTCCGTGGCGACTGAATCATATGATTTTATTTTTCAGTAGCTGTGAT 360
DB 301 AAAGTCCGTGGCGACTGAATCATATGATTTTATTTTTCAGTAGCTGTGAT 360
QY 361 GTTTTGTATTTAACTTAATCTACTCAAAATTCATAAACTCTAGACGACAAA 420
DB 361 GTTTTGTATTTAACTTAATCTACTCAAAATTCATAAACTCTAGACGACAAA 420
QY 421 CAGCTCTTCAATATGTAATAACAAAGATTTTGTAGAGCTTAAAAAGACATCCC 480
DB 421 CAGCTCTTCAATATGTAATAACAAAGATTTTGTAGAGCTTAAAAAGACATCCC 480
QY 481 ATGGAAGCTGAGTCAAGTGGGATGATCTCTCCCATTCGAATTAAGCGATTAACA 540
DB 481 ATGGAAGCTGAGTCAAGTGGGATGATCTCTCCCATTCGAATTAAGCGATTAACA 540
QY 541 CCGGCGACGACAAAGCTCCGACGTAGCCACAAACCTCTACAACTATCTGCCGACG 600
DB 541 CCGGCGACGACAAAGCTCCGACGTAGCCACAAACCTCTACAACTATCTGCCGACG 600
QY 601 AATAGGCGCGACGCTCTCTCTCCGACTTCAATTTCACTCCGATTCCTCTCTCTCC 660
DB 601 AATAGGCGCGACGCTCTCTCTCCGACTTCAATTTCACTCCGATTCCTCTCTCTCC 660
QY 661 TTGGCGACCGGACACCAACCGGACCTGCTCTGCGACCAATGATGATGCTGCC 720
DB 661 TTGGCGACCGGACACCAACCGGACCTGCTCTGCGACCAATGATGATGCTGCC 720
QY 721 GAAGCGCAGTCCCATCCCATTTGATTTTACAGAGATTAAGAGCTCAAAACATTTTC 780
DB 721 GAAGCGCAGTCCCATCCCATTTGATTTTACAGAGATTAAGAGCTCAAAACATTTTC 780
QY 781 TTAAACGATGGAATCAAGAGACATTTGAGAGCTAGGTTTGAACCGCGGCAATTCGCT 840
DB 781 TTAAACGATGGAATCAAGAGACATTTGAGAGCTAGGTTTGAACCGCGGCAATTCGCT 840
QY 841 TTGAGGACGACGCTTTAATGACCGGACAGATTCCTCAAGCTGCTGGGAAACTCTG 900
DB 841 TTGAGGACGACGCTTTAATGACCGGACAGATTCCTCAAGCTGCTGGGAAACTCTG 900
```

```
QY 901 TCTAATCTCGGCTCTAGAGAGATACATGAAGGCTCTTCTGATGATGAAGAGCTTACA 960
DB 901 TCTAATCTCGGCTCTAGAGAGATACATGAAGGCTCTTCTGATGATGAAGAGCTTACA 960
QY 961 GTATGACATGATGCTTGGGATTAAGGTAATTTGATTTGCGGAATTAATAAGTTTCTTC 1020
DB 961 GTATGACATGATGCTTGGGATTAAGGTAATTTGATTTGCGGAATTAATAAGTTTCTTC 1020
QY 1021 GTTTTAAATTTCAATGAATTTGATTAAGAGAGAACTTTTATCTAGTGAAGGTTCTCGGCGC 1080
DB 1021 GTTTTAAATTTCAATGAATTTGATTAAGAGAGAACTTTTATCTAGTGAAGGTTCTCGGCGC 1080
QY 1081 TCTCTGTGTATTCAGAGAGAGTGTGAGATGATGATGATGATGATGATGATGATGATGAT 1140
DB 1081 TCTCTGTGTATTCAGAGAGAGTGTGAGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 1141 GCTTAAGAGAGAGTGTGCTAAGTGTGTTAAGCAAGATGTGTTTATGTTATGAGCGCTGCG 1200
DB 1141 GCTTAAGAGAGAGTGTGCTAAGTGTGTTAAGCAAGATGTGTTTATGTTATGAGCGCTGCG 1200
QY 1201 GTTTCTGATGCTCGAGGAGTCTAAGGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 GTTTCTGATGCTCGAGGAGTCTAAGGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 TGAATTTGTGAGAGAGCTTTGAAGCTTTTACAGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 TGAATTTGTGAGAGAGCTTTGAAGCTTTTACAGATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 ACGAGGCTGCTTTAAGACCTTTCTGATTTGATGATGATGATGATGATGATGATGATGATGAT 1380
DB 1321 ACGAGGCTGCTTTAAGACCTTTCTGATTTGATGATGATGATGATGATGATGATGATGATGAT 1380
QY 1381 GAGAGAGAGAGCAAGTACCTTGACCGGATTTACGTGACAAATGATGATGATGATGATGATGAT 1440
DB 1381 GAGAGAGAGAGCAAGTACCTTGACCGGATTTACGTGACAAATGATGATGATGATGATGATGAT 1440
QY 1441 AGAGATCATCTCCGCTTATGCTTGAAGCTAATGCTTACCGCTTGTGATGATGATGATGATGAT 1500
DB 1441 AGAGATCATCTCCGCTTATGCTTGAAGCTAATGCTTACCGCTTGTGATGATGATGATGATGAT 1500
QY 1501 TGGGAAAGAGCTAAATGATTTAAGCGGTGCGGAAATTTTGTGCTGTGAGAGAGG 1560
DB 1501 TGGGAAAGAGCTAAATGATTTAAGCGGTGCGGAAATTTTGTGCTGTGAGAGAGG 1560
QY 1561 TGGAGCATCAGCTCTTGTGGGGGTTTGAACCCGTGAGAAATGATGATGATGATGATGATGATGAT 1620
DB 1561 TGGAGCATCAGCTCTTGTGGGGGTTTGAACCCGTGAGAAATGATGATGATGATGATGATGATGAT 1620
QY 1621 ACGAATGACAGCTGCTGAGCAGGTATACAGTTTATGATGATGATGATGATGATGATGATGATGAT 1680
DB 1621 ACGAATGACAGCTGCTGAGCAGGTATACAGTTTATGATGATGATGATGATGATGATGATGATGAT 1680
QY 1681 TGAATTAATCTTAAGTGTTCATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
DB 1681 TGAATTAATCTTAAGTGTTCATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
QY 1741 GCTAACCCAGCAATATTTCCAGAGAGCTTTGAAGTTTACAGAGTTGACATTTGCTTCTT 1800
DB 1741 GCTAACCCAGCAATATTTCCAGAGAGCTTTGAAGTTTACAGAGTTGACATTTGCTTCTT 1800
QY 1801 GTGGCTCAAGCTTTTATGTTAAGAGACCACTTTTACAGAGTCTGATTAAGCAATTC 1860
DB 1801 GTGGCTCAAGCTTTTATGTTAAGAGACCACTTTTACAGAGTCTGATTAAGCAATTC 1860
QY 1861 CAGCAATTTCAAGAGCTTAAGTAAAGGCTAATGAGGATTCGCGAGTGTGATGATGATGATGATGAT 1920
DB 1861 CAGCAATTTCAAGAGCTTAAGTAAAGGCTAATGAGGATTCGCGAGTGTGATGATGATGATGATGAT 1920
QY 1921 CCGAATTAATTTGAGAGATTAAGCTTCCGCTTGAAGAGGAGCTCTGTGACATGCTTATAGGC 1980
DB 1921 CCGAATTAATTTGAGAGATTAAGCTTCCGCTTGAAGAGGAGCTCTGTGACATGCTTATAGGC 1980
QY 1981 AAAGTTGATGAATGCCGTATGTGTTGGCTTGAACATGAGAGATTCACAAATTAAGAAAT 2040
```

Db 1981 AAAGTTAGTAAATCCGATATGTTGGGCTTAGACAGTGAAGATTCATATATAGAAAT 2040
 Qy 2041 CCAGCTATTTGGAGTTGTTTGGAGAAATTCAAATCGTATGACAAATGATATCTCCT 2100
 Db 2041 CCAGCTATTTGGAGTTGTTTGGAGAAATTCAAATCGTATGACAAATGATATCTCCT 2100
 Qy 2101 GGAATAGCAAAATTTGGGAAACCGTGGCAGGGGTTGTCCTTCCAGGTTGACAGAC 2160
 Db 2101 GGAATAGCAAAATTTGGGAAACCGTGGCAGGGGTTGTCCTTCCAGGTTGACAGAC 2160
 Qy 2161 ACCAAGATTAATAAATTTAAATCGGGGACTATATGATGATCCATGTTTGAATAC 2220
 Db 2161 ACCAAGATTAATAAATTTAACTGGGGGACTATATGATGATCCATGTTTGAATAC 2220
 Qy 2221 TTGGAAAGATGAGATGATTCAGGGTTCTCTTTAGTGTGCTGTGCAATATGGCAAG 2280
 Db 2221 TTGGAAAGATGAGATGATTCAGGGTTCTCTTTAGTGTGCTGTGCAATATGGCAAG 2280
 Qy 2281 ATTGAGCCGAGCATGTGAAGATAGTCTATGACGACGACGAAAGTTTTCCTCC 2340
 Db 2281 ATTGAGCCGAGCATGTGAAGATAGTCTATGACGACGACGAAAGTTTTCCTCC 2340
 Qy 2341 CGCTATACAGATAGAAATCTCGGCTGAACCAAGATGTGCAAGAGACAGTGTTAAGTA 2400
 Db 2341 CGCTATACAGATAGAAATCTCGGCTGAACCAAGATGTGCAAGAGACAGTGTTAAGTA 2400
 Qy 2401 GATCTGTGTGTAACAATATGAGCCGATGATGATGACCTGGTGTCTTATATGCAAGCT 2460
 Db 2401 GATCTGTGTGTAACAATATGAGCCGATGATGATGACCTGGTGTCTTATATGCAAGCT 2460
 Qy 2461 GTAGAACCTCTGAAACTTTGAAACTATATATGCAATTCAGAGCTGGGGTCTCAGAG 2520
 Db 2461 GTAGAACCTCTGAAACTTTGAAACTATATATGCAATTCAGAGCTGGGGTCTCAGAG 2520
 Qy 2521 AGTAGCGTTGATGAAACTACTGTTGAAATGTCGGTGTGATATGTTAAAGAGCAAGT 2580
 Db 2521 AGTAGCGTTGATGAAACTACTGTTGAAATGTCGGTGTGATATGTTAAAGAGCAAGT 2580
 Qy 2581 GTGAAGATCCTAGCTGTGTGAGCAATGGAATGATTCAGCTGTTCAAGCAAGATAT 2640
 Db 2581 GTGAAGATCCTAGCTGTGTGAGCAATGGAATGATTCAGCTGTTCAAGCAAGATAT 2640
 Qy 2641 TTTCTTAAAGCAAGCTCATCTTTCAAGCAAGATATGTTCTTCTATATGAAATCTGAT 2700
 Db 2641 TTTCTTAAAGCAAGCTCATCTTTCAAGCAAGATATGTTCTTCTATATGAAATCTGAT 2700
 Qy 2701 GTGCTACCAATAGTATGATTAATGATGCAATTTTCATATATCTGCAATGCTCAAAATA 2760
 Db 2701 GTGCTACCAATAGTATGATTAATGATGCAATTTTCATATATCTGCAATGCTCAAAATA 2760
 Qy 2761 TGCTTGTGTGTGAGTGAAGAAATAGTTCACCTTAATATACATATCCAAAAGTTGACC 2820
 Db 2761 TGCTTGTGTGTGAGTGAAGAAATAGTTCACCTTAATATACATATCCAAAAGTTGACC 2820
 Qy 2821 AAGATTAACAAGTGTGATGATTAATTTCACTAATATATGCTGGAATTTTGAATCAA 2880
 Db 2821 AAGATTAACAAGTGTGATGATTAATTTCACTAATATATGCTGGAATTTTGAATCAA 2880
 Qy 2881 CTGTAGACAGAAATGTAATTTCACTCTCAACATTTCTGTTAAGAAATACGTAAGATTAG 2940
 Db 2881 CTGTAGACAGAAATGTAATTTCACTCTCAACATTTCTGTTAAGAAATACGTAAGATTAG 2940
 Qy 2941 AGATGCTTATAGTGTGCTTGTCCAACTTTCTTCTTATATTTTCTTTTCGATTT 3000
 Db 2941 AGATGCTTATAGTGTGCTTGTCCAACTTTCTTCTTATATTTTCTTTTCGATTT 3000
 Qy 3001 AGGTCAGTACAGACTGACATGATGAAAGCACTTCCCAAGATGATGCTAGAACCTGACGA 3060
 Db 3001 AGGTCAGTACAGACTGACATGATGAAAGCACTTCCCAAGATGATGCTAGAACCTGACGA 3060
 Qy 3061 GAATATAGTATCCAGTGGCAGAAATTAAGTCTGTGGCTTTGGGCTGATCACCGCAT 3120
 Db 3061 GAATATAGTATCCAGTGGCAGAAATTAAGTCTGTGGCTTTGGGCTGATCACCGCAT 3120

Db 3061 GAATATAGTATCCAGTGGCAGAAATTAAGTCTGTGGCTTTGGGCTGATCACCGCAT 3120
 Qy 3121 AGAATGTTACCAAGGTGAGGGAATTAATCTCAATTCATCAATGTTGNAAACTGT 3180
 Db 3121 AGAATGTTACCAAGGTGAGGGAATTAATCTCAATTCATCAATGTTGNAAACTGT 3180
 Qy 3181 TGGAATGATTAATAGTGTGAGTGTGCTTGTGTAATCTGTTATTAATGATTTGATGGC 3240
 Db 3181 TGGAATGATTAATAGTGTGAGTGTGCTTGTGTAATCTGTTATTAATGATTTGATGGC 3240
 Qy 3241 GAATGCTGAAGATTTGACATGACAGACAGCTGAAATCTGGCAGCTTGGGTTTATG 3300
 Db 3241 GAATGCTGAAGATTTGACATGACAGACAGCTGAAATCTGGCAGCTTGGGTTTATG 3300
 Qy 3301 ATTATACATCTTAAATCTATCTGTTGACAGTGTGACAGTCTGACAGATGAAACCGTG 3360
 Db 3301 ATTATACATCTTAAATCTATCTGTTGACAGTGTGACAGTCTGACAGATGAAACCGTG 3360
 Qy 3361 CTGTGTGAAAGCAACTCTGGAGAGTCTGCTGTCTATCTGATTTGTTGATCAGAAA 3420
 Db 3361 CTGTGTGAAAGCAACTCTGGAGAGTCTGCTGTCTATCTGATTTGTTGATCAGAAA 3420
 Qy 3421 ACAATGCTACTGATGTGAGAACTTACACAAAGATAGCAAGATTTTCTGTCCAAATCAG 3480
 Db 3421 ACAATGCTACTGATGTGAGAACTTACACAAAGATAGCAAGATTTTCTGTCCAAATCAG 3480
 Qy 3481 GGTGAAATACCTGAAGGCTCTGTTGATGATATATATATATATATATATATATATATAT 3540
 Db 3481 GGTGAAATACCTGAAGGCTCTGTTGATGATATATATATATATATATATATATATATAT 3540
 Qy 3541 GAGCTTCGAGATTTCTCTTGTGTGTAATTTCTCTCTAAGTTAGTGTATTAATGA 3600
 Db 3541 GAGCTTCGAGATTTCTCTTGTGTGTAATTTCTCTCTAAGTTAGTGTATTAATGA 3600
 Qy 3601 ACACAAAAAATTAAGCTTCTGTGACACACCTTTTCTTATATCTAATATATATATATAT 3660
 Db 3601 ACACAAAAAATTAAGCTTCTGTGACACACCTTTTCTTATATCTAATATATATATATAT 3660
 Qy 3661 GCTACAA 3667
 Db 3661 GCTACAA 3667

RESULT 2
 ADJ38136
 ID ADJ38136 standard; DNA; 3667 BP.
 XX
 AC ADJ38136;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Arabidopsis thaliana AbFtn2 genomic DNA Segid10.
 XX
 KW prokaryotic type; plasmid division; Fun2; ARCE; ARCS; Fzo; plant cell;
 KW agronomic; horticultural; crop plant; ornamental plant; woody plant;
 KW herbicide target; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 OS
 PN MO2004001003-A2.
 XX
 PD 31-DEC-2003.
 XX
 PF 20-JUN-2003; 2003MO-US019536.
 XX
 PR 20-JUN-2002; 2002US-0390140P.
 PR 09-AUG-2002; 2002US-040242P.
 PR 20-JUN-2003; 2003US-00600070.
 XX
 FA (DNMS) UNIV MICHIGAN STATE.
 XX
 PI Oseferyoung KM, Vilcha S, Koksharova OA, Gao H;
 XX

DR WPI; 2004-082486/08.
DR P-PSDB; ADJ38203.
XX
PT New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
PT for further characterizing plastid division in plant cells, and in
PT varying agronomic and horticultural characteristics of economically
PT important plants.
XX
PS Claim 1; SEQ ID NO 10; 287bp; English.
XX
CC This invention relates to novel prokaryotic type or plastid division and
CC related genes and proteins, in particular, the invention relates to novel
CC Ftn2 (ARCS), ARCS and Fzo-like genes and polypeptides. The methods and
CC compositions of the present invention are useful for further
CC characterizing plastid division in plant cells, in order to vary
CC agronomic and horticultural characteristics of economically important
CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is that of a DNA sequence
CC which is related to the invention.
XX
SQ Sequence 3667 BP; 982 A; 669 C; 843 G; 1173 T; 0 U; 0 Other:
Query Match 99.9%; Score 3663.8; DB 12; Length 3667;
Best Local Similarity 99.9%; Freq. No. 0;
Matches 3665; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGTTCGATTAAGAGAAATACAAATTAAGCAATTTGCTTGAATTTCAAGAAATTTG 60
DB 1 TGTTCGATTAAGAGAAATACAAATTAAGCAATTTGCTTGAATTTCAAGAAATTTG 60
QY 61 CTGGCTAATAGGATTCATGGCTCTGTTGCTTTTACATTTACATGTCATTAATAGTTTCG 120
DB 61 CTGGCTAATAGGATTCATGGCTCTGTTGCTTTTACATTTACATGTCATTAATAGTTTCG 120
QY 121 AATTTTACATTTTCAAGTGAATGTTAAGAAAGAGGGAATTGAGGGGTTTGTGGG 180
DB 121 AATTTTACATTTTCAAGTGAATGTTAAGAAAGAGGGAATTGAGGGGTTTGTGGG 180
QY 181 TTTAACTTTTAAAGTAGTCAAGAAATTAAGTATGTTTACTGTGTCTATATGTATA 240
DB 181 TTTAACTTTTAAAGTAGTCAAGAAATTAAGTATGTTTACTGTGTCTATATGTATA 240
QY 241 AATGAAGCAACTCCAAAGGTTCTTAAAGTGAATGATTATTAAAGAGATTTAAATCAT 300
DB 241 AATGAAGCAACTCCAAAGGTTCTTAAAGTGAATGATTATTAAAGAGATTTAAATCAT 300
QY 301 AAAATCGGTGGCGACTGTAATCATATGATGTTTATTTTATTTTTCAGTGTGGAT 360
DB 301 AAAATCGGTGGCGACTGTAATCATATGATGTTTATTTTATTTTTCAGTGTGGAT 360
QY 361 GTTTTGTGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
DB 361 GTTTTGTGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
QY 421 CAGTCTCTTCAATATGTAATAAAGAAAGATTTTGTAGAGCTTAAAGAAAGACATCCC 480
DB 421 CAGTCTCTTCAATATGTAATAAAGAAAGATTTTGTAGAGCTTAAAGAAAGACATCCC 480
QY 481 ATGAAGCTCTGATCAAGTGGGCAATGTTCTCTCCCATTCGAATTAAGCCGATTAACA 540
DB 481 ATGAAGCTCTGATCAAGTGGGCAATGTTCTCTCCCATTCGAATTAAGCCGATTAACA 540
QY 541 CCGGCGAAGCAAGTCCGAGTGAAGCAAGCAAGCTTCAAGATCTGCTCCGCGAGC 600
DB 541 CCGGCGAAGCAAGTCCGAGTGAAGCAAGCAAGCTTCAAGATCTGCTCCGCGAGC 600
QY 601 AAATGGGCGGACGCTTCTCTCGACTTCAATTTCACTCCGATCTCTCTCTCTCC 660
DB 601 AAATGGGCGGACGCTTCTCTCGACTTCAATTTCACTCCGATCTCTCTCTCTCC 660
QY 661 TTGGCCACCGCGACCAACCGGCACTCTGCTCTCTGCGACCATTAATGATGCTGCC 720
DB 661 TTGGCCACCGCGACCAACCGGCACTCTGCTCTCTGCGACCATTAATGATGCTGCC 720

QY 721 GAACGCCAGTCCCATCCCATTTGATTTCTACAGGATATTAGAGCTCAAAACATTTCC 780
DB 721 GAACGCCAGTCCCATCCCATTTGATTTCTACAGGATATTAGAGCTCAAAACATTTCC 780
QY 781 TTAACCGATGATACAGAAAGCAATTTGAAAGCTTGTGAAACCGCGCAATTTGGT 840
DB 781 TTAACCGATGATACAGAAAGCAATTTGAAAGCTTGTGAAACCGCGCAATTTGGT 840
QY 841 TTGACGACGAGCTTAAATCAGCCGAGACAGATTTCTTCAAGCTGTGGAAACTCTG 900
DB 841 TTGACGACGAGCTTAAATCAGCCGAGACAGATTTCTTCAAGCTGTGGAAACTCTG 900
QY 901 TCTAATCTCGCTCTAGAAAGAGTACATGAAGTCTTGTGATGATGAAGAGCTACA 960
DB 901 TCTAATCTCGCTCTAGAAAGAGTACATGAAGTCTTGTGATGATGAAGAGCTACA 960
QY 961 GTCATCATGATGTTCTTGGGATAGGTAAATTTGATTTGGAATTAATAAGTTTCTTC 1020
DB 961 GTCATCATGATGTTCTTGGGATAGGTAAATTTGATTTGGAATTAATAAGTTTCTTC 1020
QY 1021 GTTTTAAATTCATGAATTTGAATTAAGAGAGAACTTTATCTAGTGAAGGTTCTGGGGC 1080
DB 1021 GTTTTAAATTCATGAATTTGAATTAAGAGAGAACTTTATCTAGTGAAGGTTCTGGGGC 1080
QY 1081 TCTCTGTATTTGCAAGAGTGTGAGACTGAGATAGTTCTTCGGGTTGGTGAAGCTCT 1140
DB 1081 TCTCTGTATTTGCAAGAGTGTGAGACTGAGATAGTTCTTCGGGTTGGTGAAGCTCT 1140
QY 1141 GCTTAAGAGAGTGTGCTTAAGTCTTAAAGCAAGATGCTGTTTAAATTTGATTTGGCCTTC 1200
DB 1141 GCTTAAGAGAGTGTGCTTAAGTCTTAAAGCAAGATGCTGTTTAAATTTGATTTGGCCTTC 1200
QY 1201 GTTTCTGATGTCTCGAGGAGTGTATGAGTGAATTCACATCTGATTTTAACTGTGTA 1260
DB 1201 GTTTCTGATGTCTCGAGGAGTGTATGAGTGAATTCACATCTGATTTTAACTGTGTA 1260
QY 1261 TGAATTTGTTGAAGAGCTTTGAAGCTTTTACAGATAGTTGACCTTGTGTAATTTG 1320
DB 1261 TGAATTTGTTGAAGAGCTTTGAAGCTTTTACAGATAGTTGACCTTGTGTAATTTG 1320
QY 1321 ACGAGGCTTGGCTTTAATAAGATCTTTGATTTGAATCTTTGATTTGATGCTTGTGTA 1380
DB 1321 ACGAGGCTTGGCTTTAATAAGATCTTTGATTTGAATCTTTGATTTGATGCTTGTGTA 1380
QY 1381 GAGAGAGAGCAAGTATGCTTGACACCGAATTAACGTCACAAATTAATGATGACTTTGGA 1440
DB 1381 GAGAGAGAGCAAGTATGCTTGACACCGAATTAACGTCACAAATTAATGATGACTTTGGA 1440
QY 1441 AGAGATCACTCCGCTTATGCTTGAAGCTACTTGGCTTACCGCTTGTGATGATTAACGC 1500
DB 1441 AGAGATCACTCCGCTTATGCTTGAAGCTACTTGGCTTACCGCTTGTGATGATTAACGC 1500
QY 1501 TCGGAAAAGCTAAATGTTTAAAGCGGTGTGCGAAATATTTGTGTGTGTGAGAGAG 1560
DB 1501 TCGGAAAAGCTAAATGTTTAAAGCGGTGTGCGAAATATTTGTGTGTGTGAGAGAG 1560
QY 1561 TGGAGCATGAGCTCTGTGGGGGTTTGAACCGGAGAAATTTATGAATGAGGCGTTTTT 1620
DB 1561 TGGAGCATGAGCTCTGTGGGGGTTTGAACCGGAGAAATTTATGAATGAGGCGTTTTT 1620
QY 1621 ACGAATGACAGCTGCTGAGAGAGGTATACAGTTTAACTTTTAAATTTCTTTTGA 1680
DB 1621 ACGAATGACAGCTGCTGAGAGAGGTATACAGTTTAACTTTTAAATTTCTTTTGA 1680
QY 1681 TGAATTAATCTTAAAGTTTCAATTTTAAATGATGTTGTGTGATGATCTTTTGTGA 1740
DB 1681 TGAATTAATCTTAAAGTTTCAATTTTAAATGATGTTGTGTGATGATCTTTTGTGA 1740
QY 1741 GCTAACCCAGCAATATTCAGAGAGTCAATTTGAATTTACGAAGTTGCACTTGTCTT 1800
DB 1741 GCTAACCCAGCAATATTCAGAGAGTCAATTTGAATTTACGAAGTTGCACTTGTCTT 1800

QY	1801	TTGGCTCAAGCTTTTAATTGGTAAGAAGCACACCTTTTACAGAGATGCGTAAAGCAATTC	1860
DB	1801	GTGGCTCAAGCTTTTAATTGGTAAGAAGCACACCTTTTACAGAGATGCGTAAAGCAATTC	1860
QY	1861	CAGCAACTTCACAGAGGCTAAGGTAAATGCGCTATGAGATTCCGTGCAATTTGTATATGATCA	1920
DB	1861	CAGCAACTTCACAGAGGCTAAGGTAAATGCGCTATGAGATTCCGTGCAATTTGTATATGATCA	1920
QY	1921	CGGAATAATTGGAGATATGACTTCGGTCTTAGAAGGGGACTCTGTGCACTGCTTAATAGGC	1980
DB	1921	CGGAATAATTGGAGATATGACTTCGGTCTTAGAAGGGGACTCTGTGCACTGCTTAATAGGC	1980
QY	1981	AAAGTGAATGATGCGCTAATGATGCTTGGGCTTTAGACATGAGAGATTCACATATATAGAAT	2040
DB	1981	AAAGTGAATGATGCGCTAATGATGCTTGGGCTTTAGACATGAGAGATTCACATATATAGAAT	2040
QY	2041	CCAGCTAATGAGAGTTGTTTGGAGAAATTCAAATCGTGATGACAAATGATATCTCCCT	2100
DB	2041	CCAGCTAATGAGAGTTGTTTGGAGAAATTCAAATCGTGATGACAAATGATATCTCCCT	2100
QY	2101	GGACTATGCAAAATGTTGAAAACTGTGTGGCAGGGGGTGTCTTTCTTAGGTTCAAGAC	2160
DB	2101	GGACTATGCAAAATGTTGAAAACTGTGTGGCAGGGGGTGTCTTTCTTAGGTTCAAGAC	2160
QY	2161	ACCAAAAGTAAAAAATTTAACTCGGGGACTATATGATATGATATGATATGATATGATATGATAC	2220
DB	2161	ACCAAAAGTAAAAAATTTAACTCGGGGACTATATGATATGATATGATATGATATGATATGATAC	2220
QY	2221	TTGGAAGAAGTGAAGGTAGTTCAGGGTCTCTCTTAGCTGCTGCTGCAACTATGAGCAGAG	2280
DB	2221	TTGGAAGAAGTGAAGGTAGTTCAGGGTCTCTCTTAGCTGCTGCTGCTGCAACTATGAGCAGAG	2280
QY	2281	ATTGAGCCGAGCATATGTAAGAACTATGCTATATGACAGCACTGCAGAAAGTTTTCCTTCC	2340
DB	2281	ATTGAGCCGAGCATATGTAAGAACTATGCTATATGACAGCACTGCAGAAAGTTTTCCTTCC	2340
QY	2341	CGCTATACAGATAGAAATCTCGGCTGAAACCCAGAGATGTGCAAGAGACAGTGTATATGTATA	2400
DB	2341	CGCTATACAGATAGAAATCTCGGCTGAAACCCAGAGATGTGCAAGAGACAGTGTATATGTATA	2400
QY	2401	GATCCTGTGTGGTAAACAATGTATAGGCGGTATATGCTGATATGTTTAAAGAGGCAACT	2460
DB	2401	GATCCTGTGTGGTAAACAATGTATAGGCGGTATATGCTGATATGTTTAAAGAGGCAACT	2460
QY	2461	GTAAGACCCCTCTGAAACCTTTGAAACTATATGATATATGCAATTCGAGCTGGGGTCTCAGAG	2520
DB	2461	GTAAGACCCCTCTGAAACCTTTGAAACTATATGATATATGCAATTCGAGCTGGGGTCTCAGAG	2520
QY	2521	AGTACGCTGTATGAAACTATCTGTGGAATGTTCCTGTCTGATATATGTTAAAGAGGCAACT	2580
DB	2521	AGTACGCTGTATGAAACTATCTGTGGAATGTTCCTGTCTGATATATGTTAAAGAGGCAACT	2580
QY	2581	GTAAGAGATCCATAGCTGCTGTGTGTGGCAATTTGACATGATTTCACTGCTCAGCAGAGAGAT	2640
DB	2581	GTAAGAGATCCATAGCTGCTGTGTGTGGCAATTTGACATGATTTCACTGCTCAGCAGAGAGAT	2640
QY	2641	TTTCTTAAAGCAGCTCATCTTTTCAACGCAAGAGATATGTTTCTTATGAAATCTGAT	2700
DB	2641	TTTCTTAAAGCAGCTCATCTTTTCAACGCAAGAGATATGTTTCTTATGAAATCTGAT	2700
QY	2701	GTGCGTACCAATAGGTATATTAATATGATGCAATTTTCATATATCTGCAATGCTCAAAATA	2760
DB	2701	GTGCGTACCAATAGGTATATTAATATGATGCAATTTTCATATATCTGCAATGCTCAAAATA	2760
QY	2761	TGCTGTGTTTGTAGCTAAGAACATATGTTCCACTATATATGATATGATATGATATGATATGATAC	2820
DB	2761	TGCTGTGTTTGTAGCTAAGAACATATGTTCCACTATATATGATATGATATGATATGATATGATAC	2820
QY	2821	AAAGTAAACAAGTTGCTGAGTAAATTTCACTAATATATGCTGCTGAATTTTGTATCAAA	2880
DB	2821	AAAGTAAACAAGTTGCTGAGTAAATTTCACTAATATATGCTGCTGAATTTTGTATCAAA	2880
QY	2881	CTGTATGACAGAAATGTAAATTTCACTCTCAACATTTCTGTTTAAATATACGTAGAGATTAG	2940

Dd		2881	CTGTAGACAGAAATGTAATTTCCTCAACTTTCCTGTTTAGAATAAGTAGATTAAG	2940
Oy		2941	AGATTGCCCTTAGTGTCGGCTTTGTCCAACTTTTCCTTGATTTTTTCTTTTGCATTT	3000
Dd		2941	AGATTGCCCTTAGTGTCGGCTTTGTCCAACTTTTCCTTGATTTTTTCTTTTGCATTT	3000
Oy		3001	AGGGTCAGTCAGAGCTGAACCATTTACAAGAACCCTCCAGAAATGGAATCTAGAGACTGAGA	3060
Dd		3001	AGGGTCAGTCAGAGCTGAACCATTTACAAGAACCCTCCAGAAATGGAATCTAGAGACTGAGA	3060
Oy		3061	GAAATATAGTATCCAAGTGGCAGAAAGTTAAAGTCTCTGGCTTTTGGGCGCTGATCACCGCAT	3120
Dd		3061	GAAATATAGTATCCAAGTGGCAGAAAGTTAAAGTCTCTGGCTTTTGGGCGCTGATCACCGCAT	3120
Oy		3121	AGAAATGTTACACAGAGGTGAGGGAATTAATCTAACATTCATCAATTTGTGTGAAAACGTG	3180
Dd		3121	AGAAATGTTACACAGAGGTGAGGGAATTAATCTAACATTCATCAATTTGTGTGAAAACGTG	3180
Oy		3181	TGGACATGATTAATAGTCTGGTGCCTTGTTGATTTCTGTATTTATPAGTTTGGATGGGC	3240
Dd		3181	TGGACATGATTAATAGTCTGGTGCCTTGTTGATTTCTGTATTTATPAGTTTGGATGGGC	3240
Oy		3241	GAATGCTGAAGATTGTGACCTGACAGACAGCTGAAACTGGGCGAGCTTGGGTGGTTATG	3300
Dd		3241	GAATGCTGAAGATTGTGACCTGACAGACAGCTGAAACTGGGCGAGCTTGGGTGGTTATG	3300
Oy		3301	ATTATPACACTGTTGAAACTATCTGTGTGACAGTGTGACAGTCTCAGACAGATGGAAACCCGTG	3360
Dd		3301	ATTATPACACTGTTGAAACTATCTGTGTGACAGTGTGACAGTCTCAGACAGATGGAAACCCGTG	3360
Oy		3361	CTCTGTGTGGAAGCAACTCTGTGAGGAGTCTGCTGTCTATCTGATTTGGTTCATCAGAAAA	3420
Dd		3361	CTCTGTGTGGAAGCAACTCTGTGAGGAGTCTGCTGTCTATCTGATTTGGTTCATCAGAAAA	3420
Oy		3421	ACATGCTACTGATGATGTCAGAACTCACAACAAGATACGAAGTTTCTGGCCAAAGTCAG	3480
Dd		3421	ACATGCTACTGATGATGTCAGAACTCACAACAAGATACGAAGTTTCTGGCCAAAGTCAG	3480
Oy		3481	GGTGGAAAATCACTGGAAGGCTCTGTTCTGTCATCATTAATPACTCATATGATGATGCTCT	3540
Dd		3481	GGTGGAAAATCACTGGAAGGCTCTGTTCTGTCATCATTAATPACTCATATGATGATGCTCT	3540
Oy		3541	GAGCTTCCGAGATTCTCTTTGTTCTGTAAATTTCTCTCTCAAGTTAGTGTATTAATYGA	3600
Dd		3541	GAGCTTCCGAGATTCTCTTTGTTCTGTAAATTTCTCTCTCAAGTTAGTGTATTAATYGA	3600
Oy		3601	ACACAAAAAATTTAAGCTTCTGTGSCAACCCCTTTCTGTATCTAAACCTAATPACATPAGG	3660
Dd		3601	ACACAAAAAATTTAAGCTTCTGTGSCAACCCCTTTCTGTATCTAAACCTAATPACATPAGG	3660
Oy		3661	GCTACAA 3667	
Dd		3661	GCTACAA 3667	
<hr/>				
RESULT 3				
ID	ADJ38210	standard; DNA; 2637 BP.		
XX	ADJ38210;			
XX	06-MAY-2004	(first entry)		
DE	Plastid division-related Arc6 orthologue gene 3.			
XX				
XX	prokaryotic type; plastid division; Ftn2; ARC6; ARC5; Fzo; plant cell;			
KW	agronomic; horticultural; crop plant; ornamental plant; woody plant;			
KW	herbicide target; gene; de.			
XX				
OS	Arabidopsis thaliana.			
FN	WO2004001003-A2.			

XX 31-DEC-2003.
 PD 20-JUN-2003; 2003WO-US019536.
 PF 20-JUN-2002; 2002US-0390140P.
 XX 09-AUG-2002; 2002US-0402242P.
 PR 20-JUN-2003; 2003US-00600070.
 XX (UNMS) UNIV MICHIGAN STATE.
 XX Oseeryoung KW, Vitcha S, Kokeharova OA, Gao H;
 DR WPI; 2004-082486/08.
 DR P-PSDB; ADJ38209.
 XX New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
 PT for further characterizing plastid division in plant cells, and in
 PT varying agronomic and horticultural characteristics of economically
 PT important plants.
 PS Disclosure; Fig 8; 287pp; English.
 XX This invention relates to novel prokaryotic type or plastid division and
 CC related genes and proteins. In particular, the invention relates to novel
 CC Ftn2 (AR66), ARCS and Fzo-like genes and polypeptides. The methods and
 CC compositions of the present invention are useful for further
 CC characterizing plastid division in plant cells, in order to vary
 CC agronomic and horticultural characteristics of economically important
 CC plants, such as crop, ornamental and woody plants. They can also be used
 CC as herbicide targets. The present sequence is that of a gene which is
 CC related to the invention.
 XX
 XX Sequence 2637 BP; 706 A; 535 C; 644 G; 752 T; 0 U; 0 Other;
 SQ
 Query Match 52.7%; Score 1931; DB 12; Length 2637;
 Best Local Similarity 80.4%; Pred. No. 0;
 Matches 2616; Conservative 0; Mismatches 5; Indels 632; Gaps 5;
 QY 368 GATTAACTTATCTACTCAAAATCAAAATTCATAAACCCTAGACGACAAAGTCTC 427
 DB 1 GATTAACTTATCTACTCAAAATCAAAATTCATAAACCCTAGACGACAAAGTCTC 60
 QY 428 TTCAATATGTAAACAGAAACAAGTTTTTGTAGTACCTTAAAGACATCTCCATGAG 487
 DB 61 TTCAATATGTAAACAGAAACAAGTTTTTGTAGTACCTTAAAGACATCTCCATGAG 120
 QY 488 CTCTGAGTCAAGTGGGCAATGGTCTCTCCCAATTCATTAATGCGGATTAACACGGCGA 547
 DB 121 CTCTGAGTCAAGTGGGCAATGGTCTCTCCCAATTCATTAATGCGGATTAACACGGCGA 180
 QY 548 CGAACAAGCTCCGAGTAGCAGCAACACCTTACAACTATGCTCCGCGACGAATGGG 607
 DB 181 CGAACAAGCTCCGAGTAGCAGCAACACCTTACAACTATGCTCCGCGACGAATGGG 240
 QY 608 CGAACAAGCTCTCTCGCATTTCAATTTACCTCCGATTCCTCTCTCTCTGCGCA 667
 DB 241 CGAACAAGCTCTCTCGCATTTCAATTTACCTCCGATTCCTCTCTCTCTGCGCA 300
 QY 668 CCGGACACCAACCGGCACTCTGCTCTCTGCGCAACCTATTAATGATGCTCCGAGGCC 727
 DB 301 CCGGACACCAACCGGCACTCTGCTCTCTGCGCAACCTATTAATGATGCTCCGAGGCC 360
 QY 728 ACGTCCCATCCCATTTGATTTCTACGAGTATTAAGAGCTCAAAACATTTCTTAACCG 787
 DB 361 ACGTCCCATCCCATTTGATTTCTACGAGTATTAAGAGCTCAAAACATTTCTTAACCG 420
 QY 788 ATGGAATCAGAAAGCAATTGCAAGCTAGGGTTTGAAAACCGCCCAATTGGGTTACGG 847
 DB 421 ATGGAATCAGAAAGCAATTGCAAGCTAGGGTTTGAAAACCGCCCAATTGGGTTACGG 480
 QY 848 ACGACGCTTTATATAGCCGAGACAGATTTCTCAAGCTGCTTGGGAAACTGTCTATC 907

DB 481 ACGACGCTTTATATACGCCGAGACAGATTTCTCAAGCTGCTTGGGAAACTGTCTATC 540
 QY 908 CTGCGTCTAGAAAGATAGATAGATAGATCTCTTGTAGTATGAAAGCTACATCATCA 967
 DB 541 CTGCGTCTAGAAAGATAGATAGATAGATCTCTTGTAGTATGAAAGCTACATCATCA 600
 QY 968 CTGATGTTCTTGGGATTAAGGTAATTTGCAATTCGGAATATATAAGTTCTCTGTTTAA 1027
 DB 601 CTGATGTTCTTGGGATTAAGGTAATTTGCAATTCGGAATATATAAGTTCTCTGTTTAA 617
 QY 1028 TTTCAATGAATTTGATTAAGGAAGGAACTTTATCTAGTAAGGTTCTGCGGCTCTGCT 1087
 DB 618 -----AAGTTCTGCGGCTCTGCT 638
 QY 1088 GTATTGCAAGAGGTGTGAGACTGATAGATGTTCTTGGGTTGGTGAAGCTTGTCTTAAG 1147
 DB 639 GTATTGCAAGAGGTGTGAGACTGATAGATGTTCTTGGGTTGGTGAAGCTTGTCTTAAG 698
 QY 1148 GAGAGTTGCTTAAATGCTTTAAAGCAAGATGCTTTAGTATGCGCTTGCCTTCTC 1207
 DB 699 GAGAGTTGCTTAAATGCTTTAAAGCAAGATGCTTTAGTATGCGCTTGCCTTCTC 758
 QY 1208 GATGCTCGAGGAGATGCTATGGAATGGATCCACTGATTTTATCTGTTAGATT 1267
 DB 759 GATGCTCGAGGAGATGCTATGGAATGGATCCACTGATTTTATCTGTTAGATT 818
 QY 1268 GTTGAGGAACCTTGAAGCTTTTACAGTAGTATGACTTCTTGGTATTTGAACGACG 1327
 DB 819 GTTGAGGAACCTTGAAGCTTTTACAGTAGTATGACTTCTTGGTATTTGAACGACG 843
 QY 1328 TTGGCTTTAATGAACCTTTCTTGATTTGATCTTTGTTATGAGTCTTGTAGAGGAA 1387
 DB 844 -----AGAGGAA 851
 QY 1388 GGAGCAAGTAGCCTTGACCGGATTTACGNGCAAAATGATAGACTTTGGAAGATC 1447
 DB 852 GGAGCAAGTAGCCTTGACCGGATTTACGNGCAAAATGATAGACTTTGGAAGATC 911
 QY 1448 ACTCCGCTTATGCTTGAAGTACTTGAAGCTTACCGCTTGATGATTAACGTCGAAA 1507
 DB 912 ACTCCGCTTATGCTTGAAGTACTTGAAGCTTACCGCTTGATGATTAACGTCGAAA 971
 QY 1508 AGACTAATGCTTTAGCGGTGCGGAATATTTGTGCTGTTGAGAGAGTGAAGCA 1567
 DB 972 AGACTAATGCTTTAGCGGTGCGGAATATTTGTGCTGTTGAGAGAGTGAAGCA 1031
 QY 1568 TCAAGCTTGTGGGGTTTGAACCCGTAAGATTAAGATGAGGGCTTTTACGAATG 1627
 DB 1032 TCAAGCTTGTGGGGTTTGAACCCGTAAGATTAAGATGAGGGCTTTTACGAATG 1091
 QY 1628 ACAAGCTGAGCAAGTATACAGTTTATACCTTTTAAATTTCTTTAGCATATATA 1687
 DB 1092 ACAAGCTGAGCAAGTATACAGTTTATACCTTTTAAATTTCTTTAGCATATATA 1104
 QY 1688 ACTTAAGTTTCTCATTTTAATGATGTTGTGTGTAAGTTGATCTTTTGAAGCTAACCC 1747
 DB 1105 -----AGTTGATCTTTTGTAGCTAACCC 1128
 QY 1748 CAAGCAATATTTCAAGCAAGTCAATTTGAAGTTTACGAAGTTGACCTTGTGCTC 1807
 DB 1129 CAAGCAATATTTCAAGCAAGTCAATTTGAAGTTTACGAAGTTGACCTTGTGCTC 1188
 QY 1808 AAGCTTTATTTGTAAGAACCAACACTTTTACAGATGCTGTAAGCAATTCACAGAAC 1867
 DB 1189 AAGCTTTATTTGTAAGAACCAACACTTTTACAGATGCTGTAAGCAATTCACAGAAC 1248
 QY 1868 TTACAGAGCTTAAGTATGCTATGGAATTCCTGAGATGTTGATGATACACGAAATA 1927
 DB 1249 TTACAGAGCTTAAGTATGCTATGGAATTCCTGAGATGTTGATGATACACGAAATA 1308
 QY 1928 ATTGGAGATAGACTTCGCTATGAAAAGGGGACCTGTGACAGCTTATAGGCAAGTTG 1987
 DB 1309 ATTGGAGATAGACTTCGCTATGAAAAGGGGACCTGTGACAGCTTATAGGCAAGTTG 1368

XX New isolated Ftn2, ARCS and/or Pzo-like nucleic acid sequences, useful
PT for further characterizing plasmid division in plant cells, and in
PT varying agronomic and horticultural characteristics of economically
PT important plants.

XX Claim 1; SEQ ID NO 1; 287bp; English.

XX This invention relates to novel prokaryotic type or plasmid division and
CC related genes and proteins. In particular, the invention relates to novel
CC Ftn2 (ARCS), ARCS and Pzo-like genes and polypeptides. The methods and
CC compositions of the present invention are useful for further
CC characterizing plasmid division in plant cells. In order to vary
CC agronomic and horticultural characteristics of economically important
CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is a cDNA sequence which is
CC related to the invention.

XX Sequence 2406 BP; 612 A; 492 C; 618 G; 684 T; 0 U; 0 Other;

Query Match 47.0%; Score 1724; DB 12; Length 2406;

Best Local Similarity 79.2%; Pred. No. 0;
Matches 2406; Conservative 0; Mismatches 0; Indels 632; Gaps 5;

QY 481 ATGAGAGCTTGAATCAGTCCGATTTGCTCTCCCATTCATATAGCCATTACA 540
DB 1 ATGAGAGCTTGAATCAGTCCGATTTGCTCTCCCATTCATATAGCCATTACA 60
QY 541 CCGGCGAGCAAGAGCTCCGAGCGTACGACCAACCTCTACACTATCTGCTCCGACG 600
DB 61 CCGGCGAGCAAGAGCTCCGAGCGTACGACCAACCTCTACACTATCTGCTCCGACG 120
QY 601 AAATGGGCGGACGCTCTCTCTCCGACTTCAATTTCACTCCGATTCCTCTCTCTCC 660
DB 121 AAATGGGCGGACGCTCTCTCTCCGACTTCAATTTCACTCCGATTCCTCTCTCTCC 180
QY 661 TTGGCCACCGCCACCAACCGGCACTCTGCTCTCTGCGACCACTATATGATGCC 720
DB 181 TTGGCCACCGCCACCAACCGGCACTCTGCTCTCTGCGACCACTATATGATGCC 240
QY 721 GAAGCGACGCTCCGATTTGCTCTCCGAGTATGAGAGCTCAAAACATTTTC 780
DB 241 GAAGCGACGCTCCGATTTGCTCTCCGAGTATGAGAGCTCAAAACATTTTC 300
QY 781 TTAAACGATGGAATCAGAGAGCATTCGAGAGTGGTTTGAACCGCGCAATTCGCT 840
DB 301 TTAAACGATGGAATCAGAGAGCATTCGAGAGTGGTTTGAACCGCGCAATTCGCT 360
QY 841 TTCAAGCAAGACGCTTTATATACCGCGAGACATTTCTCAAGCTGCTTGGAAACTCTG 900
DB 361 TTCAAGCAAGACGCTTTATATACCGCGAGACATTTCTCAAGCTGCTTGGAAACTCTG 420
QY 901 TCTAATCTCGGCTTAGAAGAGTATCAATGAAGTCTTCTGATGAGAGAGCTACA 960
DB 421 TCTAATCTCGGCTTAGAAGAGTATCAATGAAGTCTTCTGATGAGAGAGCTACA 480
QY 961 GTCAATCACTGATGTTCTTGGGATTAAGTATTTGATTCGAGATATATAAAGTTCTTC 1020
DB 481 GTCAATCACTGATGTTCTTGGGATTTCTTGGGATTTCTTGGGATTTCTTGGGATTT 504
QY 1021 GTTTTAATTTCAATGATGATTAAGAGAGCACTTTATCTAGTGAAGGTTCTCGGAGC 1080
DB 505 GTTTTAATTTCAATGATGATTAAGAGAGCACTTTATCTAGTGAAGGTTCTCGGAGC 518
QY 1081 TCTCTGTGATTTGCAAGAGTGTGAGACATGATAGTTCTTGGGGTGTGAGGCTCT 1140
DB 519 TCTCTGTGATTTGCAAGAGTGTGAGACATGATAGTTCTTGGGGTGTGAGGCTCT 578
QY 1141 GCTTAAAGAGAGTGTGCTTAAGTGTGTTTAAAGCAAGATGTGTTTAAAGTGTGCTTGC 1200
DB 579 GCTTAAAGAGAGTGTGCTTAAGTGTGTTTAAAGCAAGATGTGTTTAAAGTGTGCTTGC 638
QY 1201 GTTTCGATGTCTGAGAGGATGCTATGAGCATTCACCTGATTTTATCTGCTTA 1260

DB 639 GTTTCGATGTCTGAGAGGATGCTATGAGCATTCACCTGATTTTATCTGCTTA 698
QY 1261 TGAATTTTGAAGAGCTTTGAAGCTTTTACAGAGTATGATCTTCTTGTGATATTTG 1320
DB 699 TGAATTTTGAAGAGCTTTGAAGCTTTTACAGAGTATGATCTTCTTGTGATATTTG 730
QY 1321 ACAGAGCTGTGCTTATTAAGAGACTTTCTGATATTTGATATCTTGTGATATTTG 1380
DB 731 ACAGAGCTGTGCTTATTAAGAGACTTTCTGATATTTGATATCTTGTGATATTTG 731
QY 1381 GAGAGAGAGAGCAAGTACCTTGACCGGATTTACGAGCAAAATGATGAGACTTTGGA 1440
DB 732 GAGAGAGAGAGCAAGTACCTTGACCGGATTTACGAGCAAAATGATGAGACTTTGGA 791
QY 1441 AGAGATCACTCCGCTTATGCTTGGAGTACTTGGCTTACCGCTTGTGATATTTGAGC 1500
DB 792 AGAGATCACTCCGCTTATGCTTGGAGTACTTGGCTTACCGCTTGTGATATTTGAGC 851
QY 1501 TGCAGAGAGCAATGATGTTAAGCGGTGCGGATATTTTGTGCTGTGAGAGAGG 1560
DB 852 TGCAGAGAGCAATGATGTTAAGCGGTGCGGATATTTTGTGCTGTGAGAGAGG 911
QY 1561 TGAAGCACTACGCTTGTGTTGGGGTTTGAACCGGAGAAATTTATGATGAGGCTTTT 1620
DB 912 TGAAGCACTACGCTTGTGTTGGGGTTTGAACCGGAGAAATTTATGATGAGGCTTTT 971
QY 1621 ACGAATGACAGCTGCTGAGGAGTATCAGTTTATGATATCTTTTAAATTTCTTTTGA 1680
DB 972 ACGAATGACAGCTGCTGAGG----- 991
QY 1681 TGATATATCTTTAGAGTTTCTCATTTTAAATGATGTTGTTGTTGATCTTTTGTGA 1740
DB 992 TGATATATCTTTAGAGTTTCTCATTTTAAATGATGTTGTTGTTGATCTTTTGTGA 1008
QY 1741 GCTACCCCAAGCAATATTCAGAGAGTCAATTTGAAGTTTACGAAGTTGACTGCTCT 1800
DB 1009 GCTACCCCAAGCAATATTCAGAGAGTCAATTTGAAGTTTACGAAGTTGACTGCTCT 1068
QY 1801 GTGGCTCAAGCTTTTATGTTAAGAGCAACACTTTTACAGATGCTGATTAAGCAATTC 1860
DB 1069 GTGGCTCAAGCTTTTATGTTAAGAGCAACACTTTTACAGATGCTGATTAAGCAATTC 1128
QY 1861 CAGCACTTCAAGAGCTTAAGTATGCTATGAGAGTCTCTGATGTTGATATACA 1920
DB 1129 CAGCACTTCAAGAGCTTAAGTATGCTATGAGAGTCTCTGATGTTGATATACA 1188
QY 1921 CCGAATTAATTGGAGATGAGCTTCGCTGTAAGAGGAGACTCTGTGCACTGTTATAGG 1980
DB 1189 CCGAATTAATTGGAGATGAGCTTCGCTGTAAGAGGAGACTCTGTGCACTGTTATAGG 1248
QY 1981 AAAGTTGATGATGCTGATGTTGGGCTTAGACAGTGAAGATTCACATATATAGGAAT 2040
DB 1249 AAAGTTGATGATGCTGATGTTGGGCTTAGACAGTGAAGATTCACATATATAGGAAT 1308
QY 2041 CCAAGCTATTTGAGATTTGTTTGGAGATTCAAATCTGATGACATGATGATCTCCCT 2100
DB 1309 CCAAGCTATTTGAGATTTGTTTGGAGATTCAAATCTGATGACATGATGATCTCCCT 1368
QY 2101 GGAATATGAGAAATTTGTTGGAACCTGTTGGCAGGGGTTGCTTCTAGGTTCAAGAC 2160
DB 1369 GGAATATGAGAAATTTGTTGGAACCTGTTGGCAGGGGTTGCTTCTAGGTTCAAGAC 1428
QY 2161 ACCAAGATTAATAAATTTAAATCTGAGGAGCTAATGATGATCTATGTTTGAATTAC 2220
DB 1429 ACCAAGATTAATAAATTTAAATCTGAGGAGCTAATGATGATCTATGTTTGAATTAC 1488
QY 2221 TTGAGAAAGATGAGAGTATGACAGGTTTCTCTTATGCTCTGCTGCACTATGCAAG 2280
DB 1489 TTGAGAAAGATGAGAGTATGACAGGTTTCTCTTATGCTCTGCTGCACTATGCAAG 1548
QY 2281 ATTGAGCCGAGCATGTTGAAGGATGATGATGAGGCACTGAGAGAAAGTTTCTCTCC 2340

```

Db      1549 ATTGAGCCGAGCATGTGAAGTAGTGTATGACAGGACATGCAGAAAATTTCCTCC 1608
Qy      2341 CGCATACAGATAGAACTGGCTGAACCCAGAGATGCGAAGAGACAGTGTAGSTA 2400
Db      1609 CGCATACAGATAGAACTGGCTGAACCCAGAGATGCGAAGAGACAGTGTAGSTA 1668
Qy      2401 GATCTGTGTGTAACAATGTAAGGCGTGAATGTAAGGCGTGTGTCTTATTTGCAAGACT 2460
Db      1669 GATCTGTGTGTAACAATGTAAGGCGTGAATGTAAGGCGTGTGTCTTATTTGCAAGACT 1728
Qy      2461 GTAGAGCCCTGTGAAAATTGAAAATTATGATATGCAATTCAGCTGGGGCTCAGAG 2520
Db      1729 GTAGAGCCCTGTGAAAATTGAAAATTATGATATGCAATTCAGCTGGGGCTCAGAG 1788
Qy      2521 AGTAGCGTGTGAATGAACTACTGTTGAAATGTCGGTGTGATATGTAAGAGGCAAGT 2580
Db      1789 AGTAGCGTGTGAATGAACTACTGTTGAAATGTCGGTGTGATATGTAAGAGGCAAGT 1848
Qy      2581 GTGAAGATCCCTAGCTGTGTGTGTGCAATGCAATGCACTGATTTTCAGCCAGAGATAT 2640
Db      1849 GTGAAGATCCCTAGCTGTGTGTGTGCAATGCAATGCACTGATTTTCAGCCAGAGATAT 1908
Qy      2641 TTTCTTAAAAAGCACTCATCTTTTCAACGCAAGATATGTTCTTCTATGAAATCTGAT 2700
Db      1909 TTTCTTAAAAAGCACTCATCTTTTCAACGCAAGATATGTTCTTCTATGAAATCTGAT 1968
Qy      2701 GTGCTACCCATAGTATGATTAATGATGCAATTTTCATATATGCAATTCCTCAATA 2760
Db      1969 GTGCTACCA----- 1978
Qy      2761 TGCTGTGTTGTGAGTAAAGAACATAGTCCCACTAATACATGTCCCAAAAGTTGATC 2820
Db      1979 ----- 1978
Qy      2821 AAGATTAACAAGTTCGTGATTAATTTCACTAATATATGCTGTAATTTTGTGATCAA 2880
Db      1979 ----- 1978
Qy      2881 CTGTAGACAGAAATGTAATTTCACTCATCAATTTCTGTTAGAAATACAGATTAG 2940
Db      1979 ----- 1978
Qy      2941 AGATGCTTAGTGTGCTTGTTCCACTTTCTTCTGATTTTCTTTTCAATTT 3000
Db      1979 ----- 1978
Qy      3001 AGGCTCAGTCAGAGCTGACGATTCAGAGCACTTCCAGAAATGATGCTAGACTGCAGA 3060
Db      1980 AGGCTCAGTCAGAGCTGACGATTCAGAGCACTTCCAGAAATGATGCTAGACTGCAGA 2039
Qy      3061 GAATATGATATCCAGTGGCGAAGATTAATCTCTGGCTTTGGGCTGTATCACCGCAT 3120
Db      2040 GAATATGATATCCAGTGGCGAAGATTAATCTCTGGCTTTGGGCTGTATCACCGCAT 2099
Qy      3121 AGAAATGTTACAGAGGTGAGGAATAATCTCAATCAATGATTTGTGAAGAACTGT 3180
Db      2100 AGAAATGTTACAG----- 2113
Qy      3181 TGACATGATTAATAGTGTGTGCTTGTGATTTCTGTTAATTAAGTTTGTGATGGC 3240
Db      2114 -----AGGTTTGTGATGGC 2128
Qy      3241 GAATGCTGAAGATTTGATGTCAGACAGAGCTGAAAATCGCGCACTTGGTGTGTTATG 3300
Db      2129 GAATGCTGAAGATTTGATGTCAGACAGAGCTGAAAATCGCGCACTTGGTGTGTTATG 2188
Qy      3301 ATTATACACTGTTGAATCTATCTGTGACAGTGTGACAGCTTCACGACAGATGAACCCGTG 3360
Db      2189 ATTATACACTGTTGAATCTATCTGTGACAGTGTGACAGCTTCACGACAGATGAACCCGTG 2248
Qy      3361 CTCTGTGTGAAGCAACTCTGAGAGAGTCTGCTTGTATCTGATTTGGTTGATCAGAAA 3420
Db      2249 CTCTGTGTGAAGCAACTCTGAGAGAGTCTGCTTGTATCTGATTTGGTTGATCAGAAA 2308

```

```

Qy      3421 ACAATGCTACTGATGTGAGAACTACACAACAAGATGAAGTTTCTGTGTCGAATGAG 3480
Db      2309 ACAATGCTACTGATGTGAGAACTACACAACAAGATGAAGTTTCTGTGTCGAATGAG 2368
Qy      3481 GGTGAAAAATCACTGAAGCTCTGTCTTTCGATCATATA 3518
Db      2369 GGTGAAAAATCACTGAAGCTCTGTCTTTCGATCATATA 2406

RESULT 5
ADJ38135
ID      ADJ38135 standard; cDNA; 2406 BP.
XX
XX
AC      ADJ38135;
XX
DT      06-MAY-2004 (first entry)
XX
DE      Arabidopsis thaliana AtFtn2 cDNA SegID9.
XX
KW      prokaryotic type; plastid division; Ftn2; ARC6; ARC5; Fzo; plant cell;
KW      agronomic; horticultural; crop plant; ornamental plant; woody plant;
KW      herbicide target; gene; ss.
XX
OS      Arabidopsis thaliana.
XX
PN      WO2004001003-A2.
XX
PD      31-DEC-2003.
XX
PF      20-JUN-2003; 2003WO-US019536.
XX
PP      20-JUN-2002; 2002US-0390140P.
PR      09-AUG-2002; 2002US-040242P.
PR      20-JUN-2003; 2003US-0060070.
XX
PA      (UNMS ) UNIV MICHIGAN STATE.
PI
PI      Osheryoung KW, Vicha S, Koksharova OA, Gao H;
XX
XX      WP1; 2004-082486/08.
DR      P-PSDB; ADJ38203.
XX
PT      New isolated Ftn2, ARC5 and/or Fzo-like nucleic acid sequences, useful
PT      for further characterizing plastid division in plant cells, and in
PT      varying agronomic and horticultural characteristics of economically
PT      important plants.
XX
XX
PS      Claim 1; SEQ ID NO 9; 287bp; English.
XX
CC      This invention relates to novel prokaryotic type or plastid division and
CC      related genes and proteins. In particular, the invention relates to novel
CC      Ftn2 (ARC6), ARC5 and Fzo-like genes and polypeptides. The methods and
CC      compositions of the present invention are useful for further
CC      characterizing plastid division in plant cells, in order to vary
CC      agronomic and horticultural characteristics of economically important
CC      plants, such as crop, ornamental and woody plants. They can also be used
CC      as herbicide targets. The present sequence is a cDNA sequence which is
CC      related to the invention.
XX
SQ      Sequence 2406 BP; 611 A; 491 C; 619 G; 685 T; 0 U; 0 Other;

Query Match      46.9%; Score 1720.8; DB 12; Length 2406;
Beet Local Similarity 79.1%; Pred. No. 0;
Matches 2404; Conservative 0; Mismatches 2; Indels 632; Gaps 5;

Qy      481 ATGGAACCTGAGTCACTGAGCTGATGATGCTCTCCCAATTCATATATGCGGATTAACA 540
Db      1 ATGGAACCTGAGTCACTGAGCTGAGCTGATGATGCTCTCCCAATTCATATATGCGGATTAACA 60

Qy      541 CCGGCGAGCAAAAGCTCCGAGCTGAGCACAACAACCTCTACAACTATCTGCTCCGACG 600
Db      61 CCGGCGAGCAAAAGCTCCGAGCTGAGCACAACAACCTCTACAACTATCTGCTCCGACG 120

```

QY 601 AAATGGGCGGACGCTTCTCTCCGACTTCAATTTCACCTCCGATTCTCTCTCTCC 660
|
|
|
Db 121 AAATGGGCGGACGCTTCTCTCCGACTTCAATTTCACCTCCGATTCTCTCTCTCC 180
|
|
|
QY 661 TTGGCCACCGGACACCAACCGGCACTCTGCTCTCTGCGCAACATTAATTGATGCTCC 720
|
|
|
Db 181 TTGGCCACCGGACACCAACCGGCACTCTGCTCTCTGCGCAACATTAATTGATGCTCC 240
|
|
|
QY 721 GAAGCGGACGCGGACATCCCATTTGATTCTTCAAGGATTAAGAGCTCAAAACATTTTC 780
|
|
|
Db 241 GAAGCGGACGCGGACATCCCATTTGATTCTTCAAGGATTAAGAGCTCAAAACATTTTC 300
|
|
|
QY 781 TTAACCGATGGAATCAGAAAGAGATTGGAAGCTAGGGTTTCGAAAACCGCGCAATTCGCT 840
|
|
|
Db 301 TTAAACGATGGAATCAGAAAGAGATTGGAAGCTAGGGTTTCGAAAACCGCGCAATTCGCT 360
|
|
|
QY 841 TTCAGCGACGACGCTTTAATCAAGCGGAGACAGATTCTTCAAGCTGCTTCCGAAACTCTG 900
|
|
|
Db 361 TTCAGCGACGACGCTTTAATCAAGCGGAGACAGATTCTTCAAGCTGCTTCCGAAACTCTG 420
|
|
|
QY 901 TCTAATCTGCTGTAGAAAGAGATCAATGAAGTCTTCTGATGATGAAGAGCTTACA 960
|
|
|
Db 421 TCTAATCTGCTGTAGAAAGAGATCAATGAAGTCTTCTGATGATGAAGAGCTTACA 480
|
|
|
QY 961 GTCATCACTGATGTTCTTGGGATTAAGGTAATTTTCGAAATTAATAAGTTTCTTC 1020
|
|
|
Db 481 GTATCACTGATGTTCTTGGGAT----- 504
|
|
|
QY 1021 GTTTTAATTTCAATATGCAATTAAGAGAACTTTTATCTAGTGAAGTTCTTGGGAC 1080
|
|
|
Db 505 -----AAGTTCTTGGGAC 518
|
|
|
QY 1081 TCTCTGATATTCGAAAGAGTGGTGAAGTCTGATGATGTTCTTGGGTTGGTGAAGCTCT 1140
|
|
|
Db 519 TCTCTGATATTCGAAAGAGTGGTGAAGTCTGATGATGTTCTTGGGTTGGTGAAGCTCT 578
|
|
|
QY 1141 GCTTAAAGAGAGTGGTCTAAGTCTTTTAAAGCAAGATGTGGTTTAAAGTAAAGCGCTTGC 1200
|
|
|
Db 579 GCTTAAAGAGAGTGGTCTAAGTCTTTTAAAGCAAGATGTGGTTTAAAGTAAAGCGCTTGC 638
|
|
|
QY 1201 GTTTCCTGATGTCGAGAGGATGCTAATGCAATTCGATTCACCTGATTTTATTAATCTGCTTA 1260
|
|
|
Db 639 GTTTCCTGATGTCGAGAGGATGCTAATGCAATTCGATTCACCTGATTTTATTAATCTGCTTA 698
|
|
|
QY 1261 TGAAGTTGTGAGAGCTTGAAGCTTTTACAGTAAGTTGACTGTGCTTGTGAATTTG 1320
|
|
|
Db 699 TGAAGTTGTGAGAGCTTGAAGCTTTTAC----- 730
|
|
|
QY 1321 ACGAGCGTGGCTTTAAGAACTTCTGTGATTTGATACCTTTGTATTTGAGTCTTGTGTA 1380
|
|
|
Db 731 -----A 731
|
|
|
QY 1381 GAGAGAAAGAGCAAGTACCTTTCACCGGATTTACGTGCACAAATTTGATGAGCTTTGGA 1440
|
|
|
Db 732 GAGAGAAAGAGCAAGTACCTTTCACCGGATTTACGTGCACAAATTTGATGAGCTTTGGA 791
|
|
|
QY 1441 AGAGATACCTCGGCTTATGCTTGAAGCTACTTGGCTTACCGGTTGGTGAATTAAGC 1500
|
|
|
Db 792 AGAGATACCTCGGCTTATGCTTGAAGCTACTTGGCTTACCGGTTGGTGAATTAAGC 851
|
|
|
QY 1501 TGGCAAAAAGCTAAATGTTTAAAGCGGTGTCGGAATATTTTGTGCTGTGTGAGAGAG 1560
|
|
|
Db 852 TGGCAAAAAGCTAAATGTTTAAAGCGGTGTCGGAATATTTTGTGCTGTGTGAGAGAG 911
|
|
|
QY 1561 TGGAGCATCAAGCTTTTGTGGGGGTTTGACCCGTGAGAAAGTTTATGAATGAGGGGTTTTT 1620
|
|
|
Db 912 TGGAGCATCAAGCTTTTGTGGGGGTTTGACCCGTGAGAAAGTTTATGAATGAGGGGTTTTT 971
|
|
|
QY 1621 ACGAATACAGCTGCTGAGAGATATACAGTTTATGATACCTTTTAAATTTCTTTGACA 1680
|
|
|
Db 972 ATGATATACAGCTGCTGAGC----- 991
|
|
|

QY 1681 TGAATTAATTTAGGTTTCTCAATTTAATGATGTTGTGGTGAAGTTGATCTTTTGTGA 1740
|
|
|
Db 992 -----AGGTTGATCTTTTGTGA 1008
|
|
|
QY 1741 GCTAACCCAGCAATATTCAGAGAGTCAATTGAAGTTTACGAAGTTGACTTGCCTT 1800
|
|
|
Db 1009 GCTAACCCAGCAATATTCAGAGAGTCAATTGAAGTTTACGAAGTTGACTTGCCTT 1068
|
|
|
QY 1801 GTGGCTCAAGCTTTTATGTTAGAAAGCCACACTTTTACAGATGCTGATTAAGCAATTC 1860
|
|
|
Db 1069 GTGGCTCAAGCTTTTATGTTAGAAAGCCACACTTTTACAGATGCTGATTAAGCAATTC 1128
|
|
|
QY 1861 CAGCACTTCAAGAGCTTAAGTATGCTATGAGATTCCTGCAATGTTGATGATACA 1920
|
|
|
Db 1129 CAGCACTTCAAGAGCTTAAGTATGCTATGAGATTCCTGCAATGTTGATGATACA 1188
|
|
|
QY 1921 CGGAATTAATTGGAGATGAGCTTCGGCTGTGAAGGGGACTCTGTGCAAGCTTATAGGC 1980
|
|
|
Db 1189 CGGAATTAATTGGAGATGAGCTTCGGCTGTGAAGGGGACTCTGTGCAAGCTTATAGGC 1248
|
|
|
QY 1981 AAAGTTGATGAATGCCGATATGCTGTTGGGCTTGAACAGTGAAGATTCAATATAGGAAT 2040
|
|
|
Db 1249 AAAGTTGATGAATGCCGATATGCTGTTGGGCTTGAACAGTGAAGATTCAATATAGGAAT 1308
|
|
|
QY 2041 CCAAGCTATTTGAGAGTTTGTTTGAGAGATTCAATTCGATGACAAATGATGATCTCCCT 2100
|
|
|
Db 1309 CCAAGCTATTTGAGAGTTTGTTTGAGAGATTCAATTCGATGACAAATGATGATCTCCCT 1368
|
|
|
QY 2101 GGAATATGCAATTTGTTGGAACCTGGTTGGCAGGGGTTGCTTTCTCTAGATTCAAGAC 2160
|
|
|
Db 1369 GGAATATGCAATTTGTTGGAACCTGGTTGGCAGGGGTTGCTTTCTCTAGATTCAAGAC 1428
|
|
|
QY 2161 ACCAATGATTAATAATTTTAACTCGGGGACTACTATGATGATCTATGATTTGATTAAC 2220
|
|
|
Db 1429 ACCAATGATTAATAATTTTAACTCGGGGACTACTATGATGATCTATGATTTGATTAAC 1488
|
|
|
QY 2221 TTGAAAAGAGTGAAGTGAAGTCTTCAAGGTTCTCTTAACTGCTGCTGCAACTATGAGCAAG 2280
|
|
|
Db 1489 TTGAAAAGAGTGAAGTGAAGTCTTCAAGGTTCTCTTAACTGCTGCTGCAACTATGAGCAAG 1548
|
|
|
QY 2281 ATTGAGCCGAGGATGGAAGGCTAAGTCAAGGCACTGAGCAAAAGTTTTCCTTCC 2340
|
|
|
Db 1549 ATTGAGCCGAGGATGGAAGGCTAAGTCAAGGCACTGAGCAAAAGTTTTCCTTCC 1608
|
|
|
QY 2341 CGCTATACAGATGAAGAACTCGGCTGAACCCAGAGATGTGCAAGACAGTGTTAAGTGA 2400
|
|
|
Db 1609 CGCTATACAGATGAAGAACTCGGCTGAACCCAGAGATGTGCAAGACAGTGTTAAGTGA 1668
|
|
|
QY 2401 GATCTGTTGTTAACTATGAGCCCGTATGATGTTGAGCTGTGTTATTTGCAAGAACT 2460
|
|
|
Db 1669 GATCTGTTGTTAACTATGAGCCCGTATGATGTTGAGCTGTGTTATTTGCAAGAACT 1728
|
|
|
QY 2461 GTAAGACCCCTCTGAAAACCTTGAACCTAATGATTAATGCAATTTGAGCTGGGGTCTGAG 2520
|
|
|
Db 1729 GTAAGACCCCTCTGAAAACCTTGAACCTAATGATTAATGCAATTTGAGCTGGGGTCTGAG 1788
|
|
|
QY 2521 AGTAGCGTTGATGAACCTACTGTTGAAGATGCTGCTGCTGATATGTTAAAGAGGCAAGT 2580
|
|
|
Db 1789 AGTAGCGTTGATGAACCTACTGTTGAAGATGCTGCTGCTGATATGTTAAAGAGGCAAGT 1848
|
|
|
QY 2581 GTGAAGATCTTACGCTGCTGCTGCTGCTGCAATTTGACTGATTTCACTGTTCAAGCAAGTAT 2640
|
|
|
Db 1849 GTGAAGATCTTACGCTGCTGCTGCTGCTGCAATTTGACTGATTTCACTGTTCAAGCAAGTAT 1908
|
|
|
QY 2641 TTTCTTAAAGCAGCTCATCTTTTCAAGCAAGGATATGAGTTCTTATGGAATCGAT 2700
|
|
|
Db 1909 TTTCTTAAAGCAGCTCATCTTTTCAAGCAAGGATATGAGTTCTTATGGAATCGAT 1968
|
|
|
QY 2701 GTGCTACCAATGAGTATGATTAATGATGCAATTTTCAATATCTGATGCTCAAAATA 2760
|
|
|
Db 1969 GTGCTACCA----- 1978
|
|
|
QY 2761 TGCTTGTTTTGTGAGCTAAGAACATAGTCTCCACTTAATATCAATGTCCCAAAAGTTGTACC 2820
|
|
|


```

Db 1979 ----- 1978
QY 2821 AAGATTACAGATTGCTGAGTAATTTCACTAATTAGTCTGGAATTTTTCATCAAA 2880
Db 1979 ----- 1978
QY 2881 CTGTAGACAGAAATGTAATTTCACTCTCAACATTTCTGTTAGAAATAGTAGATTAG 2940
Db 1979 ----- 1978
QY 2941 AGATTGCTTAGTGTGCTTTGTCACATTTCTTCTTGATTTTCTTTTCGATTT 3000
Db 1979 ----- 1979
QY 3001 AGGCTCAGTCAGAGCTGACGATTCAGAAAGCACTTCCAGAAATGATGCTAGAGCTCAGA 3060
Db 1980 AGGCTCAGTCAGAGCTGACGATTCAGAAAGCACTTCCAGAAATGATGCTAGAGCTCAGA 2039
QY 3061 GAATATAGTATCCAGAGTGGCGAGATTAATGCTCTGGCTTTTGGGCTGATCCCGCAT 3120
Db 2040 GAATATAGTATCCAGAGTGGCGAGATTAATGCTCTGGCTTTTGGGCTGATCCCGCAT 2099
QY 3121 AGAAATGTTACAGAGGTGAGGGAATTAATCTCAATTCATCAATTGTTGAAACTGT 3180
Db 2100 AGAAATGTTACAG----- 2113
QY 3181 TGGACATGATTAATAGTCTGCTGCTTTGATTTCTGTTAATTAAAGTTTGGATGGGC 3240
Db 2114 -----AGGTTTGGATGGGC 2128
QY 3241 GAATGCTGAAGATTTTGGACTGACAGAGAGCTGAACCTGGCAGCTTGGGTTGTTATG 3300
Db 2129 GAATGCTGAAGATTTTGGACTGACAGAGAGCTGAACCTGGCAGCTTGGGTTGTTATG 2188
QY 3301 ATTATACACTGTTGAAACTATCTGTTGACAGTGTGACAGTCTCAGAGATGAAACCCGTG 3360
Db 2189 ATTATACACTGTTGAAACTATCTGTTGACAGTGTGACAGTCTCAGAGATGAAACCCGTG 2248
QY 3361 CTCTGTGGAAGCACTCTGTGAGAGAGTCTGTTCTATCTGATTTGTTGTTTCATCCAGAA 3420
Db 2249 CTCTGTGGAAGCACTCTGTGAGAGAGTCTGTTCTATCTGATTTGTTGTTTCATCCAGAA 2308
QY 3421 ACATGCTACTGATGTCAGAACTTACACACAAAGATAGAAATTTCTGTGCTCAAGTCAG 3480
Db 2309 ACATGCTACTGATGTCAGAACTTACACACAAAGATAGAAATTTCTGTGCTCAAGTCAG 2368
QY 3481 GGTGAAATCACTGAAGGCTCTGTTCTTTCATCATTA 3518
Db 2369 GGTGAAATCACTGAAGGCTCTGTTCTTTCATCATTA 2406

```

```

XX 20-JUN-2002; 2002US-0390140P.
PR 09-AUG-2002; 2002US-0402242P.
PR 20-JUN-2003; 2003US-00600070.
XX
XX (UNMS ) UNIV MICHIGAN STATE.
XX
PI Oseveryoung KW, Vilha S, Koksharova OA, Gao H;
XX
DR WPI, 2004-082486/08.
DR P-PSDB; ADJ38207.
XX
PT New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
PT for further characterizing plasmid division in plant cells, and in
PT varying agronomic and horticultural characteristics of economically
PT important plants.
XX
PS Disclosure; Fig 8; 287pp; English.
XX
CC This invention relates to novel prokaryotic type or plasmid division and
CC related genes and proteins. In particular, the invention relates to novel
CC Ftn2 (ARCS), ARCS and Fzo-like genes and polypeptides. The methods and
CC compositions of the present invention are useful for further
CC characterizing plasmid division in plant cells, in order to vary
CC agronomic and horticultural characteristics of economically important
CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is that of a gene which is
CC related to the invention.
XX
SQ Sequence 2406 BP; 612 A; 493 C; 618 G; 683 T; 0 U; 0 Other;

Query Match      46.8%; Score 1717.6; DB 12; Length 2406;
Best Local Similarity 79.1%; Pred. No. 0;
Matches 2402; Conservative 0; Mismatches 4; Indels 632; Gaps 5;

QY 481 ATGGAAGCTGTAGTCAAGTGGGATTTGTTCTCCCATTCGAATTATGCGGATTAACA 540
Db 1 ATGGAAGCTGTAGTCAAGTGGGATTTGTTCTCCCATTCGAATTATGCGGATTAACA 60
QY 541 CCGGCGACGACAAAGCTCCGACGTCAGGACCAACACCTCTTACACTATCTGTCGGCCAGC 600
Db 61 CCGGCGACGACAAAGCTCCGACGTCAGGACCAACACCTCTTACACTATCTGTCGGCCAGC 120
QY 601 AAATGGGCGGACGCTTCTCTCCGACTTCAATTGCACTCCGATTCCTCTCTCTCC 660
Db 121 AAATGGGCGGACGCTTCTCTCCGACTTCAATTGCACTCCGATTCCTCTCTCTCC 180
QY 661 TTCGCAACCGGACACACACGCACTCTGTCCTCTGCGCACCATCTATTTGATCGTCC 720
Db 181 TTCGCAACCGGACACACACGCACTCTGTCCTCTGCGCACCATCTATTTGATCGTCC 240
QY 721 GAACGCCACGTCGCCATCCCATTTGATTTCTTACAGGATTTAGAGCTCAAAACATTTTC 780
Db 241 GAACGCCACGTCGCCATCCCATTTGATTTCTTACAGGATTTAGAGCTCAAAACATTTTC 300
QY 781 TTAAACGATGGAATCAAGAGCAATTGAGGCTGATTTGAAACCGCGCAATTCGGT 840
Db 301 TTAAACGATGGAATCAAGAGCAATTGAGGCTGATTTGAAACCGCGCAATTCGGT 360
QY 841 TTCAGCAGCAGCGCTTTAATCAGCCGAGACAGATTTCTTCAAGCTGCTTGGAAACTCTG 900
Db 361 TTCAGCAGCAGCGCTTTAATCAGCCGAGACAGATTTCTTCAAGCTGCTTGGAAACTCTG 420
QY 901 TCTAATCTCGGCTTAAAGAGAGTACATGAAAGTCTTCTTATGATGAAAGAGCTTCA 960
Db 421 TCTAATCTCGGCTTAAAGAGAGTACATGAAAGTCTTCTTATGATGAAAGAGCTTCA 480
QY 961 GTCATCACTGATGTTCTTGGGATTAAGTAAATTCGATTTGGAATATAAAGTTTCTTC 1020
Db 481 GTCATCACTGATGTTCTTGGGAT----- 504
QY 1021 GTTTAATTTCAATGATGATTAAGAGAGAACTTTATCTAGTGAAGTTCTCGGGCC 1080

```

```

RESULT 6
ADJ38208
ID ADJ38208 standard; DNA; 2406 BP.
AC ADJ38208;
DT 06-MAY-2004 (first entry)
XX
XX Plasmid division-related Arc6 orthologue gene 2.
DE
XX
XX prokaryotic type; plasmid division; Ftn2; ARCS; Fzo; plant cell;
XX agronomic; horticultural; crop plant; ornamental plant; woody plant;
XX herbicide target; gene; ds.
OS Arabidopsis thaliana.
XX
XX W02004001003-A2.
XX
XX 31-DEC-2003.
XX
XX 20-JUN-2003; 2003WO-US019536.
PF

```

Db 505 -----AAGTTCCTGTC 518
Qy 1081 TCTCTGTATTTGCAAGAGGTGTGAGACTGAGATAGTCTTCTGGGGTGGTAGGCTCT 1140
Db 519 TCTCTGTATTTGCAAGAGGTGTGAGACTGAGATAGTCTTCTGGGGTGGTAGGCTCT 578
Qy 1141 GCTTAAGAGAGGTGTGCTTAAGTGTGTTAAGCAAGATGTGTTTTAGTTATAGCGCTTGC 1200
Db 579 GCTTAAGAGAGGTGTGCTTAAGTGTGTTAAGCAAGATGTGTTTTAGTTATAGCGCTTGC 638
Qy 1201 GTTTCGAGTGTCTCGAGGGATGCTATGAGATTGGATCCACCTGATTTTATATCTGTTA 1260
Db 639 GTTTCGAGTGTCTCGAGGGATGCTATGAGATTGGATCCACCTGATTTTATATCTGTTA 698
Qy 1261 TGAGTTGTGAGAGACTTGAAGCTTTTACAGATAGTTGACTTGTGTTAATTTG 1320
Db 699 TGAGTTGTGAGAGACTTGAAGCTTTTAC----- 730
Qy 1321 ACGAGCGTGGCTTATATAGAACTTTGATTTGATTACTTTGTTATGAGCTTGTGTA 1380
Db 731 -----A 731
Qy 1381 GAGAGAGAGAGCAAGTAGCTTGCACCGGATTTACGTGCAAAATTTGATGAGACTTTGGA 1440
Db 732 GAGAGAGAGAGCAAGTAGCTTGCACCGGATTTACGTGCAAAATTTGATGAGACTTTGGA 791
Qy 1441 AGAGATCACTCCGCGTATGTCTTGAGACTTGGCTTACCCTGCTGATGATTAAGC 1500
Db 792 AGAGATCACTCCGCGTATGTCTTGAGACTTGGCTTACCCTGCTGATGATTAAGC 851
Qy 1501 TGGCAAAAGACTAATGTTTAAAGCGGTGTGCGGAATATTTTGTGTCTGTGAGAGG 1560
Db 852 TGGCAAAAGACTAATGTTTAAAGCGGTGTGCGGAATATTTTGTGTCTGTGAGAGG 911
Qy 1561 TGAAGACTCAAGCTCTTGTGGGGGTTGAACCCGTGAAGATTTATGATAGAGGCTTTT 1620
Db 912 TGAAGACTCAAGCTCTTGTGGGGGTTGAACCCGTGAAGATTTATGATAGAGGCTTTT 971
Qy 1621 ACGAATGACAGCTGTGAGCAGGTATCAGTTTATGATTAACCTTTTAAATTTCTTTAGCA 1680
Db 972 ACGAATGACAGCTGTGAGC----- 991
Qy 1681 TGAATATATCTTAGGTTCTCATTTTATATGATGTGTGAGGTTGATCTTTTGTGA 1740
Db 992 -----AGTTGATCTTTTGTGA 1008
Qy 1741 GCTACCCCAAGCAATATTCAGACGAGTCAATTTGAATTTACGAAGTTGACCTTGCCTT 1800
Db 1009 GCTACCCCAAGCAATATTCAGACGAGTCAATTTGAATTTACGAAGTTGACCTTGCCTT 1068
Qy 1801 GTGGCTCAAGCTTTATTTGGTAAAGAACCAACCTTTTACAGATGCTGATTAAGCAATTC 1860
Db 1069 GTGGCTCAAGCTTTATTTGGTAAAGAACCAACCTTTTACAGATGCTGATTAAGCAATTC 1128
Qy 1861 CAGCAACTTACAGAGCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1920
Db 1129 CAGCAACTTACAGAGCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 1188
Qy 1921 CCGAATTAATTTGGAAGATAGACTTCGTCTAGAAAGGGGACCTGTGTGACCTGTTAAGGC 1980
Db 1189 CCGAATTAATTTGGAAGATAGACTTCGTCTAGAAAGGGGACCTGTGTGACCTGTTAAGGC 1248
Qy 1981 AAAGTTGATGATAGCCGTATGTGGTGTGAGACTTGAACAGTGAAGATTAATAGAAAT 2040
Db 1249 AAAGTTGATGATAGCCGTATGTGGTGTGAGACTTGAACAGTGAAGATTAATAGAAAT 1308
Qy 2041 CCAAGTATTTGTGAGTTGTTTGTGAGAAATTCAAATGTGATGCAATGATGATCTCCT 2100
Db 1309 CCAAGTATTTGTGAGTTGTTTGTGAGAAATTCAAATGTGATGCAATGATGATCTCCT 1368
Qy 2101 GGAATATGCAAAATTTGTGAGAACTGTGTGAGAGGGGTGTCTTTCTTAGTTCAAGAC 2160
Db 1369 GGAATATGCAAAATTTGTGAGAACTGTGTGAGAGGGGTGTCTTTCTTAGTTCAAGAC 1428

Qy 2161 ACCAAGATTAATAATTTAAACTCGGGGACTACTATGATGATCTATGTTTGAATTAC 2220
Db 1429 ACCAAGATTAATAATTTAAACTCGGGGACTACTATGATGATCTATGTTTGAATTAC 1488
Qy 2221 TTGGAAGAAGTGAAGGTATGTTCAAGGTTCTCTTTAGTGTCTGTCTGCAACTATAGCAAG 2280
Db 1489 TTGGAAGAAGTGAAGGTATGTTCAAGGTTCTCTTTAGTGTCTGTCTGCAACTATAGCAAG 1548
Qy 2281 ATTGAGCCGAGCATGTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2340
Db 1549 ATTGAGCCGAGCATGTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1608
Qy 2341 CGCTATACAGATTAAGAACTCGGCTGAACCCAGAGATGTGCAAGACAGTGTTAAGTGA 2400
Db 1609 CGCTATACAGATTAAGAACTCGGCTGAACCCAGAGATGTGCAAGACAGAGTGTTAAGTGA 1668
Qy 2401 GATCTGTGTGTGAACAATGATAGCCGTGATGTGAGCTGTGTCTTTATTTGCAAGAGCT 2460
Db 1669 GATCTGTGTGTGAACAATGATAGCCGTGATGTGAGCTGTGTCTTTATTTGCAAGAGCT 1728
Qy 2461 GTAAAGCCCTCTGAAAACTTTGAAACTAATGATTAATGCAATTCGAGCTGGGCTCTGAG 2520
Db 1729 GTAAAGCCCTCTGAAAACTTTGAAACTAATGATTAATGCAATTCGAGCTGGGCTCTGAG 1788
Qy 2521 AGTAGCGTTGATGAAACTACGTGTGAATGTCCGTTGCTGATATGTTAAAGAGGCAAGT 2580
Db 1789 AGTAGCGTTGATGAAACTACGTGTGAATGTCCGTTGCTGATATGTTAAAGAGGCAAGT 1848
Qy 2581 GTGAAGTCTTACTGTGTGTGTGAGCAATTTGATGATGATGATGATGATGATGATGATGAT 2640
Db 1849 GTGAAGTCTTACTGTGTGTGTGAGCAATTTGATGATGATGATGATGATGATGATGATGAT 1908
Qy 2641 TTTCTTAAAGAGCTCATCTTTTCAACGCAAGATATGTTCTTCTATGATGATGAT 2700
Db 1909 TTTCTTAAAGAGCTCATCTTTTCAACGCAAGATATGTTCTTCTATGATGATGAT 1968
Qy 2701 GTGCTACCAATAGATATGATTAATGATGATGATGATGATGATGATGATGATGATGAT 2760
Db 1969 GTGCTACCA----- 1978
Qy 2761 TGCTTGTGTGTGAGCTTAAGACATAGTCCCACTTAATGATGCCAAAAGTTGACC 2820
Db 1979 ----- 1978
Qy 2821 AAGATTAACAAGTGTGAGTAATTTCACTAATTAATGCTGTTGAATTTTGTGATCAAA 2880
Db 1979 ----- 1978
Qy 2881 CTGTAGACAGAAATGTAAATTTGACTCTCAACATTTCTGTGTAGAAATACGTAGAGTTAG 2940
Db 1979 ----- 1978
Qy 2941 AGATTGCTTGAATGTGAGCTTGTGCAACTTTTCTTCTGATTTTCTTTTCCATTT 3000
Db 1979 -----T 1979
Qy 3001 AGGTCAGTCAAGCTGACGATTCAGAGCACTTCCAGAAATGATGATGATGATGATGATGAT 3060
Db 1980 AGGTCAGTCAAGCTGACGATTCAGAGCACTTCCAGAAATGATGATGATGATGATGATGAT 2039
Qy 3061 GAATATAGTATCCAAAGTGTGAGAAATTAAGTCTGTGCTTTTGGGCTGTGATCACCGCAT 3120
Db 2040 GAATATAGTATCCAAAGTGTGAGAAATTAAGTCTGTGCTTTTGGGCTGTGATCACCGCAT 2099
Qy 3121 AGAAATGTTACCAAGAGTGAAGGAATTAATCTCAATTAATTAATTTGTGTGAAGACTGT 3180
Db 2100 AGAAATGTTACCAAG----- 2113
Qy 3181 TGAACATGATTAATGATGTGAGTGTGCTTGTGATTTCTGTTATTAATGATTTGATGGG 3240
Db 2114 -----AGTTTGTGATGGG 2128

Ds 1509 TGCTGCTATTGCAAAACTTGCTGC 1532

RESULT 11

ACN48781
ID ACN48781 standard; cDNA; 607 BP.

AC ACN48781;

DT 02-DEC-2004 (first entry)

DE Cotton primed seed EST Clone ID: LIB3825-027-Q6-K6-H1, SEQ:3562.

XX Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;
XX variety DP50B; library LIB3825; molecular tag; molecular marker;
KM genetic mapping; molecular mapping; seed germination; plant growth;
XX plant quality; plant yield; plant breeding; tissue printing; ss.

OS Gossypium hirsutum.

XX US2004123340-A1.

XX 24-JUN-2004.

XX 12-DEC-2001; 2001US-00021323.

XX 14-DEC-2000; 2000US-0255619P.

XX (DEIK/) DEIKMAN J.

XX (FENG/) FENG P C C.

XX (FINC/) FINCHER K L.

XX (ZIEG/) ZIEGLER T E.

XX Delkman J, Feng PCC, Fincher KL, Ziegler TE;

XX WPI; 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its

XX fragment, useful for isolating a variety of agronomically significant

XX PT genes associated with plant growth, quality or yield, and as molecular

XX PT tags to map genes.

XX Claim 1; SEQ ID NO 3562; 34pp; English.

XX The invention relates to 17880 cotton expressed sequence tags (ESTs;

XX ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated

XX from primed or non-primed seeds from variety DP50B, mature seeds from

XX variety Coker 312, Boswell 96 Field, and androecium tissue, gynoecium

XX tissue, developing fibres, carpel walls and septa from variety

XX Nicotiana3B. The invention also relates to substantially purified

XX proteins or their fragments encoded by nucleic acid molecules of the

XX invention, and to transformed plants having a nucleic acid construct

XX comprising a nucleic acid of the invention. The cotton ESTs are useful as

XX molecular tags to isolate genetic regions, to isolate genes, to map

XX genes, to determine gene function and to determining whether genes are

XX members of a particular gene family. The nucleic acid molecules may be

XX used for isolating a variety of agronomically significant genes

XX associated with plant growth, quality, yield, and could also serve as

XX links in metabolic and catabolic pathways. The nucleic acid molecules are

XX also useful for identifying genes important in initiating and maintaining

XX seed germination or that may be used to mitigate stresses encountered

XX during seed germination. The ESTs additionally enable the acquisition of

XX promoters and cis-regulatory elements which will be useful to express

XX agronomically significant genes in these tissues and/or other tissues.

XX and also permits the acquisition of molecular markers useful in breeding

XX schemes, genetic and molecular mapping, and in cloning of agronomically

XX significant genes. The nucleic acid molecules are further useful for

XX detecting the expression level or pattern of a protein or mRNA and for

XX detecting the presence or quantity of a protein by tissue printing. The

XX present sequence represents a specifically claimed EST isolated from a

XX cotton variety DP50B primed seed cDNA library (LIB3825). The sequence

XX data for this patent did not form part of the printed specification, but

CC was obtained in electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?docID=US20040123340
XX
XX Sequence 607 BP; 172 A; 103 C; 164 G; 167 T; 0 U; 1 Other;

Query Match 4.9%; Score 178.6; DB 13; Length 607;
Best Local Similarity 69.3%; Pred. No. 1.1e-35;
Matches 266; Conservative 0; Mismatches 100; Indels 18; Gaps 1;

Ds 1986 TGATGAATGCCGATGATGTTGGGCTTACAGACAGTGAAGATTCACAAATATAGAAATCCAGC 2045

Db 5 TGNAGAGTCCCGTTCGTCGTTGGGCTTACAGACAGTGAAGATTCACAAATATAGAAATATATC 64

Qy 2046 TATTGTGAGATTTGTTTGGAGAAATTCATCAATCGATGACATGATGATTCCTCTGACT 2105

Db 65 TATTGTGAAATTTGTTTGGAGAAATTCATCAATCGATGACATGATGATTCCTCTGACT 124

Qy 2106 ATGCAAATTTGTTGAAACCTGTTGGACAGGGGTTGCTTCCATGATTCAGACACACCA 2165

Db 125 TTGCMAACTGCTGAGAGCATGGCTTAATGAGAGTGTTTCCATGATTTAGACACCA 184

Qy 2166 AGATTAATAATTTAACTCGGGAGCTACTGATGATTCCTATGATTTTGGATTAAGTGA 2225

Db 185 AGATTAATCAATTTCAAGCTTGAAGATTTATATGATTCCTACTGCTCTGATATTTAGA 244

Qy 2226 AAGATGAGAGTATGTTCAAGGTTCTCTTAACTGCTGCTGCAACTATGCAAGATTTGG 2285

Db 245 AAGCTTGAAGAGAGAGAGTGTTCACCCCTTGCTGCAAGCGGAGCTATATGAGATTTAG 304

Qy 2286 AGCCGAG-----CATGGAAGAGTATGCTATGCAAGCAAGCACTGACAGA 2327

Db 305 TGCTGAGGACTACTGACAGTTCTTATGATCAATGTAAGGCTATGCAATTTCAAGCATTTGAGAA 364

Qy 2328 AGTTTTCCTTCCCGCTATACAGA 2351

Db 365 GGTGTTTCTTCTTCTGTCGTCAGA 388

RESULT 12

ADP93143

ID ADP93143 standard; cDNA; 439 BP.

XX ADP93143;

XX 09-SEP-2004 (first entry)

XX Cotton expressed sequence tag, EST, #2154.

XX Cotton; ss; EST; expressed sequence tag; plant; plant protection;

XX plant improvement; marker-assisted breeding.

XX Gossypium hirsutum; variety Nuocotton33B.

XX US2004123338-A1.

XX 24-JUN-2004.

XX 08-DEC-2000; 2000US-00732627.

XX 10-DEC-1999; 99US-0170255P.

XX (FINC/) FINCHER K L.

XX Fincher KL;

XX WPI; 2004-479807/45.

XX New substantially purified nucleic acid molecule that encodes a cotton

XX PT protein or its fragment, useful as molecular tool for the targeting and

XX PT isolation of novel genes for plant protection and improvement.

XX Claim 1; SEQ ID NO 2154; 30pp; English.

CC The invention relates to a substantially purified nucleic acid molecule
CC that encodes a cotton protein or its fragment comprising an EST
CC (expressed sequence tag) appearing as ADP90990-ADP95919. Also included by a
CC are a substantially purified cotton protein or its fragment encoded by a
CC nucleic acid molecule above and a transformed plant (having a nucleic
CC acid molecule which comprises: an exogenous promoter region which
CC functions in a plant cell to cause the production of a mRNA molecule; a
CC structural nucleic acid molecule comprising one of the ESTs or their
CC complements; a 3' non-translated sequence that functions in the plant
CC cell to cause termination of transcription and addition of polyadenylated
CC ribonucleotides to a 3' end of the mRNA molecule). The ESTs are useful as
CC molecular tool for the targeting and isolation of novel genes for plant
CC protection and improvement. The ESTs are useful for developing new
CC strategies for understanding critical plant developmental and metabolic
CC pathways, for isolating genes and promoters, for identifying and mapping
CC the genes involved in developmental and metabolic pathways, and for
CC determining gene function. The cotton nucleic acid molecules are useful
CC as molecular tags to isolate genetic regions, isolate genes, map genes,
CC and determine gene function. The nucleic acid molecules are useful for
CC determining if genes are members of a particular gene family and for use
CC in marker-assisted breeding programs. The present sequence is one of the
CC 4330 cotton ESTs of the invention. NOTE: The sequences are not displayed
CC in the specification but are available in electronic format from the
CC USPTO at seqdata.uspto.gov/sequence.html?DocID=20040123338.
XX
SQ Sequence 439 BP; 87 A; 153 C; 79 G; 119 T; 0 U; 1 Other;
Query Match 4.8%; Score 175; DB 12; Length 439;
Best Local Similarity 65.0%; Pred. No. 8.1e-35;
Matches 278; Conservative 0; Mismatches 141; Indels 9; Gaps 1;
QY 499 GTGGGCAATGGTCTCTCCCATTCACATGACGAGCGGAGCAAAAGCTC 558
DB 12 GTAGCAATCGGGCTGTGACATCCGACTTGTCTTATTCACCCGTGAACCTCTC 71
QY 559 CGAGTGAACCAACACCTCTAACAATATCTGCTCCGCGCAAAATGAGCGAGCTCTT 618
DB 72 AACTCCACACCGTTCACTACGCTGCTGCTCCGCAATTAATGCGCGAGCGCTC 131
QY 619 CTCTCGCATTTCAATTTCACTCTG-----ATTCTCTCTCTCTCTCTTGGCAAC 669
DB 132 CTGTGCTGATCTTCAATTTCTCTCCGCGCGCAATTAATTCCTCTCTCTCTCTCA 191
QY 670 GCCACGACCAACCGGACCTGTCTCTGTCGACATCTATTGATCGCGGAGCGAC 729
DB 192 ACCGCACTCTTTTCTCCCTTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 251
QY 730 GTCCCATTCCTTGAATTTCTTACAGGTATTAGAGCTCAACACATTTCTTAACGAT 789
DB 252 GTTTCATTCCTCTTGAATTTCTTACAGGTATTAGAGCTCAACATTTCTTAAGTAT 311
QY 790 GGAATCAGAAAGAGATTCGAAAGCTGTTGAAACCGCGCAATTCGTTTACGAC 849
DB 312 GGAATTTGAAGAGCTTATGAGCAAGGGTTTGAACCGCTCAATATGAGGTTCACTCA 371
QY 850 GAGCTTTAATCAGCGGAGACATTTCTTGAAGCTGTTGCGCAACTCTCTTAATCT 909
DB 372 GACACCAATTAATTCAGCAAGACATTTCTTGAAGCTGTTGAAACCTTAATCACTT 431
QY 910 CGGCTAG 917
DB 432 GGGCTTAG 439

RESULT 13

ADJ38255 ID ADJ38255 standard; cDNA; 545 BP.

XX AC ADJ38255;

XX DT 06-MAY-2004 (first entry)

XX DE Plasmid division-related Arc6 orthologue cDNA 42.

XX prokaryotic type; plasmid division; Ftn2; ARC6; ARC5; Fzo; plant cell;
KW agronomic; horticultural; crop plant; ornamental plant; woody plant;
KM herbicide target; gene; ss.
OS *Gossypium arboreum*.
XX MO2004001003-A2.
XX 31-DEC-2003.
XX 20-JUN-2003; 2003WO-US019536.
XX 20-JUN-2002; 2002US-0390140P.
PR 09-AUG-2002; 2002US-0402242P.
PR 20-JUN-2003; 2003US-0060070.
XX (UNMS) UNIV MICHIGAN STATE.
XX
PI Osteeryoung KW, Vicha S, Kokeharova OA, Gao H;
XX WPI; 2004-082486/08.
XX
XX New isolated Ftn2, ARC5 and/or Fzo-like nucleic acid sequences, useful
PT for further characterizing plasmid division in plant cells, and in
PT varying agronomic and horticultural characteristics of economically
PT important plants.
XX
XX Disclosure; Fig 8; 287pp; English.
XX
XX This invention relates to novel prokaryotic type or plasmid division and
CC related genes and proteins. In particular, the invention relates to novel
CC Ftn2 (ARC6), ARC5 and Fzo-like genes and polypeptides. The methods and
CC compositions of the present invention are useful for further
CC characterizing plasmid division in plant cells, in order to vary
CC agronomic and horticultural characteristics of economically important
CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is a cDNA sequence which is
CC related to the invention.
XX
SQ Sequence 545 BP; 164 A; 107 C; 120 G; 151 T; 0 U; 3 Other;
Query Match 4.4%; Score 161; DB 12; Length 545;
Best Local Similarity 69.6%; Pred. No. 4e-31;
Matches 218; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
QY 3224 ATAGTTTGGATGGGCAATGCTGAAGATTGAGCTGACAGACGACCTGAACCTGCGCA 3283
DB 82 AGAGTTCTGGATGGTCAATGTTGAAGACATGAGACGATGTCAGCGCAATCGCTCA 141
QY 3284 GCTTGGGTTGTTATGATTATACACTGTGGAACCTATCTGTTGACAGTGTGACAGTCTC 3343
DB 142 GCTTGGTGGTAAATTAATATGCTTACGAAACATGCCAATGACAGTGTACCTTTC 201
QY 3344 AGCAGATGAACCCGCTCTGTTGGAAGCACTGAGAGAGCTGCTGTCTATCTGA 3403
DB 202 ACTGATGGCGAGGAGCTGATGCAAGCTTCTGGAAGATTCACCTCTTGACTGA 261
QY 3404 TTGTTTCATTCAGAAAACATGCTACTGATGTGCAACCTACACAAAGATACGAAGT 3463
DB 262 TGTTCATCATCCGAGAAACATGCTCTAATGTAACTCTTACACACGAGATATGAGAT 321
QY 3464 TTTCTGTTCAAGTCAAGGTTGGAATAATCACTGAAGCTCTGTTTGTGATCATATATAC 3523
DB 322 GTCTTGTTCACACTCAGAGCTGGAATAATCACTGAAGATCTGTCTAACAAATCTTAATGAT 381
QY 3524 TCATATGATGAT 3536
DB 382 ATGTATTAAGCAT 394

RESULT 14

ACN48855/c

ID ACN48855 standard; cDNA; 552 BP.
 XX ACN48855;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Cotton primed seed EST Clone ID: LIB3825-027-06-N6-H1, SEQ:3636.
 XX
 KM Cotton; plant; EST; expressed sequence tag; transgenic plant; seed;
 KM variety DP50B; library LIB3825; molecular tag; molecular marker;
 KM genetic mapping; molecular mapping; seed germination; plant growth;
 KM plant quality; plant yield; plant breeding; tissue printing; ss.
 XX
 OS *Gossypium hirsutum*.
 XX
 UN US2004123340-A1.
 XX
 PD 24-JUN-2004.
 XX
 PF 12-DEC-2001; 2001US-00021323.
 XX
 PR 14-DEC-2000; 2000US-0255619P.
 XX
 PA (DEIK/) DEIKMAN J.
 PA (FENG/) FENG P C C.
 PA (FINC/) FINCHER K L.
 PA (ZIEG/) ZIEGLER T E.
 XX
 PI Delkman J, Feng PCC, Fincher KL, Ziegler TE;
 DR WPI; 2004-479808/45.
 XX
 PT New isolated nucleic acid molecule that encodes a plant protein or its
 PT fragment, useful for isolating a variety of agronomically significant
 PT genes associated with plant growth, quality or yield, and as molecular
 PT tags to map genes.
 XX
 PS Claim 1; SEQ ID NO 3636; 34pp; English.
 CC
 CC The invention relates to 17880 cotton expressed sequence tags (ESTs;
 CC ACN43220-ACN63099). The ESTs were isolated from cDNA libraries generated
 CC from primed or non-primed seeds from variety DP50B, mature seeds from
 CC variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
 CC tissue, developing fibres, carpel walls and septa from variety
 CC Ncotton33B. The invention also relates to substantially purified
 CC proteins or their fragments encoded by nucleic acid molecules of the
 CC invention, and to transformed plants having a nucleic acid construct
 CC comprising a nucleic acid of the invention. The cotton ESTs are useful as
 CC molecular tags to isolate genetic regions, to isolate genes, to map
 CC genes, to determine gene function and to determine whether genes are
 CC members of a particular gene family. The nucleic acid molecules may be
 CC used for isolating a variety of agronomically significant genes
 CC associated with plant growth, quality, yield, and could also serve as
 CC links in metabolic and catabolic pathways. The nucleic acid molecules are
 CC also useful for identifying genes important in initiating and maintaining
 CC seed germination or that may be used to mitigate stresses encountered
 CC during seed germination. The ESTs additionally enable the acquisition of
 CC promoters and cis-regulatory elements which will be useful to express
 CC agronomically significant genes in these tissues and/or other tissues.
 CC and also permits the acquisition of molecular markers useful in breeding
 CC schemes, genetic and molecular mapping, and in cloning of agronomically
 CC significant genes. The nucleic acid molecules are further useful for
 CC detecting the expression level or pattern of a protein or mRNA and for
 CC detecting the presence or quantity of a protein by tissue printing. The
 CC present sequence represents a specifically claimed EST isolated from a
 CC cotton variety DP50B primed seed cDNA library (LIB3825). The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from the US patent office at
 CC seqdata.uspo.gov/sequence.html?docid=US20040123340
 XX
 SEQ Sequence 552 BP; 153 A; 122 C; 110 G; 167 T; 0 U; 0 Other;
 Query Match 4.3%; Score 159.4; DB 13; Length 552;

Best Local Similarity 69.3%; Pred. No. 1.1e-30;
 Matches 217; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
 QY 3224 ATAGTTTGGATGAGCGGAATGCTGAAGATTGGACTGACAGACAGCTGAACTGCGCA 3283
 DB 444 AGAGGTTTCGATGCTGAATGTTGAAGACATGACAGCATGCGACGCAAAATCGCTCA 385
 QY 3284 GCTTGGGTTGTTTATGATTATACACTGTTGAAACTATCTGTTGACAGTGTACAGTCTC 3343
 DB 384 GCTTGGTGGTATATGAAATATGTTACTACGAACATGCGCATTTGACAGTGTACCCCTTC 325
 QY 3344 AGCAGATGGAACCCGCTCTGTTGGAAGCACTGAGAGAGTCTGCTATCTGA 3403
 DB 324 ACTAGATGSCCAGCGAGCTTATGTCAGAGTACTTGAAGAAATCCACTGCTTGACTGA 265
 QY 3404 TTTGGTTCATCCAGAAACCAATCTACTGATGTGACAACTGACAAAGATACGAGT 3463
 DB 264 TGTTCATCATCCGAGAACAAATGCTTATGTATTAATCTTACACGAGATATGATGAT 205
 QY 3464 TTTCTGTGCAAGTCAAGGTGGAATAATCACTGAAGGCTGTTCTTGATCATATATATAC 3523
 DB 204 GTCTTGTTCACACTCAGGCTGGAATAATCACTGAAGGATCTGTACAAATCTTAGTATG 145
 QY 3524 TCATATGTAGCAT 3536
 DB 144 ATGTATTAAGCAT 132
 RESULT 15
 ADJ38217
 ID ADJ38217 standard; cDNA; 608 BP.
 XX
 AC ADJ38217;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Plastid division-related Arce. orthologue cDNA 10.
 XX
 KM prokaryotic type; plastid division; Ftn2; ARCE; ARCS; Fzo; plant cell;
 KM agronomic; horticultural; crop plant; ornamental plant; woody plant;
 KM herbicide target; gene; ss.
 XX
 OS *Solanum tuberosum*.
 XX
 UN WO2004001003-A2.
 XX
 PD 31-DEC-2003.
 XX
 PF 20-JUN-2003; 2003WO-US019536.
 XX
 PR 20-JUN-2002; 2002US-0390140P.
 PR 09-AUG-2002; 2002US-0402242P.
 PR 20-JUN-2003; 2003US-00600070.
 XX
 PA (UNMS) UNIV MICHIGAN STATE.
 XX
 PI OeserYoung KW, Vitsha S, Koksharova OA, Gao H;
 DR WPI; 2004-082466/08.
 XX
 PT New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
 PT for further characterizing plastid division in plant cells, and in
 PT varying agronomic and horticultural characteristics of economically
 PT important plants.
 XX
 PS Disclosure; Fig 8; 287bp; English.
 XX
 CC This invention relates to novel prokaryotic type or plastid division and
 CC related genes and proteins. In particular, the invention relates to novel
 CC Ftn2 (ARCS), ARCS and Fzo-like genes and polypeptides. The methods and
 CC compositions of the present invention are useful for further
 CC characterizing plastid division in plant cells, in order to vary
 CC agronomic and horticultural characteristics of economically important

CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is a cDNA sequence which is
CC related to the invention.

XX
SQ Sequence 608 BP; 146 A; 167 C; 128 G; 167 T; 0 U; 0 Other;

Query Match 4.3%; Score 158; DB 12; Length 608;

Best Local Similarity 63.2%; Pred. No. 2.6e-30;

Matches 264; Conservative 0; Mismatches 145; Indels 9; Gaps 1;

```
QY 581 CAACTATCTGCTCCGACGAAATGGGCGACCGCTCTTCTCCGACTTCAATT----- 635
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 CTACTAATCTCTCCGCTAGTAATGGGCGATCGTCTTCGCGATTTCCAATTCCTTC 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 636 ----CACCTCGATTCCTCCCTCCCTCTCTCTCGCCACCGCACCGCCCACTCTCG 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 CTTCACACACACACCTCCGACTCATCGATTCGAAATTCAGAAATTCACAACTCCGTTA 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 TCTCTGCGACACATCTATGATGCTCCGAAACGACAGTCCCATCCCATTTGATTCT 751
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 CGACTATTCCTCCTCCTGTTGCTCTCTTCAGACACACACATTTCAATGCTATAGACTTTT 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 752 ACCAGTATTAGAGCTCAACACATTTCTTAACCGATGAATCAAGAAAGACATTGAA 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 ATAGAGTGTGTGTGAAGCTCACTTCCTCGGTAGCGTATTAGAGATGTACGATG 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 812 CTAGGGTTTGAACCCGCGAATTCCGTTTCAGCGAGAGGCTTTAATCAGCCGAGAC 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 CTAGAAATTACAAAGCCCTCCGAGTACGATACAGTCAAGAAAGCAATGTGGCCGACGGC 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 872 AGATTCTTCAAGCTGTGCGAAACTGTCTAATCTCGGCTTGAAGAGATACAATG 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 AGATTCTTCAAGCTGTGCGAAACCTTGCTGACTTCACTCTCTGTAAGAGTACAATC 528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 932 AAGGCTCTTGTGATGATGAAGAGCTACAGTCACTGATGTTCTTGGGATTAAGT 989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 AAGGCTCTGCTCAGCATGATGTTGATTAATTTCTAATCTCTGCTCCCTGGGATTAAGT 586
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: June 9, 2005, 17:27:57
Job time : 1826 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2005, 17:28:08 ; Search time 94 Seconds
(without alignments)
3295.692 Million cell updates/sec

Title: US-10-600-070-2

Perfect score: 4063

Sequence: 1 MEALSHVIGISLPQLCRP.....YEVFWSKSGMKITGSLAS 801

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4063	100.0	801	ADJ38202	Adj38202 Arabidops
2	4051	99.7	801	ADJ38207	Adj38207 plastid d
3	4051	99.7	801	ADJ38211	Adj38211 plastid d
4	4051	99.7	801	ADJ38209	Adj38209 plastid d
5	1775.5	43.7	760	ADJ38205	Adj38205 plastid d
6	1646	40.5	324	ADJ38203	Adj38203 Arabidops
7	524.5	12.9	768	ADJ38247	Adj38247 plastid d
8	498.5	12.3	789	ADJ38274	Adj38274 plastid d
9	497	12.2	798	ADJ38244	Adj38244 plastid d
10	497	12.2	798	ADJ38245	Adj38245 plastid d
11	442	10.9	631	ADJ38242	Adj38242 plastid d
12	442	10.9	631	ADJ38132	Adj38132 Synchoco
13	442	10.9	631	ADJ38240	Adj38240 plastid d
14	431	10.6	819	ADJ38253	Adj38253 plastid d
15	431	10.6	819	ADJ38251	Adj38251 plastid d
16	429	10.6	714	ADJ38249	Adj38249 plastid d
17	429	10.6	714	ADJ38250	Adj38250 plastid d
18	423	10.4	673	ADJ38272	Adj38272 plastid d
19	413	10.2	836	ADJ38270	Adj38270 plastid d
20	401	9.9	716	ADJ38239	Adj38239 plastid d
21	344.5	8.5	702	ADJ38235	Adj38235 plastid d
22	320	7.9	661	ADJ38237	Adj38237 plastid d
23	145.5	3.6	6095	AAU10701	Amino aci
24	143	3.5	2089	AAW08333	Cyclolell
25	139.5	3.4	954	AAW72752	DNA polym

26	139.5	3.4	954	2	ADJ87484	Adj87484 Confectio
27	138	3.3	2021	7	ADJ70511	Adj70511 Human hea
28	133.5	3.3	6973	4	ADJ26982	Adj26982 Sorangium
29	133	3.3	1273	4	ABG00113	ABG00113 Novel hum
30	132.5	3.3	876	8	ADM13791	ADM13791 Bacillus
31	131.5	3.2	2111	4	AAJ66471	AAJ66471 Protein e
32	130.5	3.2	876	6	ABJ43335	ABJ43335 Bacillus
33	130	3.2	1082	5	ABP62883	ABP62883 Human pol
34	128.5	3.2	876	2	AAJ80137	AAJ80137 A. steatoc
35	128.5	3.2	876	2	AAJ80143	AAJ80143 B. steatoc
36	128.5	3.2	876	2	AAJ80142	AAJ80142 B. steatoc
37	128	3.2	9234	7	ADJ26980	Adj26980 Sorangium
38	127.5	3.1	2472	4	AAU33568	AAU33568 Pseudomon
39	127	3.1	1241	3	AAJ93337	AAJ93337 A human i
40	127	3.1	1241	6	ADA14856	ADA14856 Human int
41	127	3.1	1241	7	ABR42349	ABR42349 Human int
42	127	3.1	1241	8	ADP46554	ADP46554 Cancer-as
43	126.5	3.1	1169	5	ABP29745	ABP29745 Streptoco
44	126.5	3.1	1224	5	ABP26662	ABP26662 Streptoco
45	126.5	3.1	2482	7	ABO68221	ABO68221 Pseudomon

ALIGNMENTS

RESULT 1	ADJ38202	ADJ38202 standard; protein; 801 AA.
ID	ADJ38202;	
XX	ADJ38202;	
AC		
XX		
DT	06-MAY-2004	(first entry)
XX		
DE	Arabidopsis thaliana Arc6-1 protein.	
XX		
KW	prokaryotic type; plastid division; Ftn2; ARC6; ARC5; Fzo; plant cell; agronomic; horticultural; crop plant; ornamental plant; woody plant; herbicide target.	
KW		
XX		
OS	Arabidopsis thaliana.	
XX		
PN	WO2004001003-A2.	
XX		
PD	31-DEC-2003.	
XX		
PF	20-JUN-2003; 2003WO-US019536.	
XX		
PR	20-JUN-2002; 2002US-0390140P.	
PR	09-AUG-2002; 2002US-040242P.	
PR	20-JUN-2003; 2003US-00600070.	
XX		
PA	(UNMS) UNIV MICHIGAN STATE.	
PI	Osteryoung KW, Vitla S, Kosharova OA, Gao H;	
XX		
XX	WPI: 2004-082486/08.	
DR	N-PSDB; ADJ38129; ADJ38130.	
XX		
PT	New isolated Ftn2, ARC5 and/or Fzo-like nucleic acid sequences, useful for further characterizing plastid division in plant cells, and in varying agronomic and horticultural characteristics of economically important plants.	
PT		
XX		
PS	Disclosure; Fig 2; 287bp; English.	
XX		
CC	This invention relates to novel prokaryotic type or plastid division and related genes and proteins. In particular, the invention relates to novel Ftn2 (ARC6), ARC5 and Fzo-like genes and polypeptides. The methods and compositions of the present invention are useful for further characterizing plastid division in plant cells, in order to vary agronomic and horticultural characteristics of economically important plants, such as crop, ornamental and woody plants. They can also be used as herbicide targets. The present sequence is that of a protein which is	
CC		

CC related to the invention.

XX Sequence 801 AA;

Query Match 100.0%; Score 4063; DB 8; Length 801;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MEALSHVIGIGLSPFQOLCRLPATTKLRSHNTSTTICSAKMDRLSDPNFTSDSSSS 60
DB 1 MEALSHVIGIGLSPFQOLCRLPATTKLRSHNTSTTICSAKMDRLSDPNFTSDSSSS 60
QY 61 PATATTATVLSLPPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFG 120
DB 61 PATATTATVLSLPPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDDALISRQIIOACETLSNPRSREYNEGILLDDEATVITDVPMDKVPALCVLOEG 180
DB 121 FSDDALISRQIIOACETLSNPRSREYNEGILLDDEATVITDVPMDKVPALCVLOEG 180
QY 181 GETEIVLRVGEALLKERLPKSFQODVVLVMAALFLDVSRDAMALDPPDFTGYEFVEAL 240
DB 181 GETEIVLRVGEALLKERLPKSFQODVVLVMAALFLDVSRDAMALDPPDFTGYEFVEAL 240
QY 241 KLIQEGASSLAPDLRAQIDETLEITPRVYLLGLPLGDDVAAKRLNGLSGVRNLTWS 300
DB 241 KLIQEGASSLAPDLRAQIDETLEITPRVYLLGLPLGDDVAAKRLNGLSGVRNLTWS 300
QY 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVEVALVAQA 360
DB 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVEVALVAQA 360
QY 361 FIGKPHLLQDADKQFQOLQQAQKVMAMEIPAMLYDTRNWEIDGLERGLCALLIGKVD 420
DB 361 FIGKPHLLQDADKQFQOLQQAQKVMAMEIPAMLYDTRNWEIDGLERGLCALLIGKVD 420
QY 421 CRMWLGIDSEDSQYRNPAIVEFVLENSNRDNDLPGICLLFTWLAGVFPFRDTRDK 480
DB 421 CRMWLGIDSEDSQYRNPAIVEFVLENSNRDNDLPGICLLFTWLAGVFPFRDTRDK 480
QY 481 KFKLGDYDDPMVLSYERVEVVGSPLAATAATMARI GAEHVKASAMQALQKVPSTRTD 540
DB 481 KFKLGDYDDPMVLSYERVEVVGSPLAATAATMARI GAEHVKASAMQALQKVPSTRTD 540
QY 541 RNSAEPDQVETVSVDPVGNNGRDGEPGFIAEAVRPSSENFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPDQVETVSVDPVGNNGRDGEPGFIAEAVRPSSENFETNDYAIRAGVSESSVD 600
QY 601 ETTVEKSVADMKEASVKILAAVAILGLISLFSQKTFIKSSSSFQKQDMSMESDVATI 660
DB 601 ETTVEKSVADMKEASVKILAAVAILGLISLFSQKTFIKSSSSFQKQDMSMESDVATI 660
QY 661 GSVRADSESLPRMDATTAENIVSKWKIKSLAGPDRILEMLPEVLDGRLKLTWRBAA 720
DB 661 GSVRADSESLPRMDATTAENIVSKWKIKSLAGPDRILEMLPEVLDGRLKLTWRBAA 720
QY 721 ETAQGLAVYDTLLKLKSVDSVTVSADGTALVEATLESACLSDLVHPENNATDVRTYTT 780
DB 721 ETAQGLAVYDTLLKLKSVDSVTVSADGTALVEATLESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFMSKGMKITEGSLAS 801
DB 781 RYEVFMSKGMKITEGSLAS 801

```

RESULT 2

ID ADJ38207 standard; protein; 801 AA.

XX AC ADJ38207;

XX DT 06-MAY-2004 (first entry)

DE Plastid division-related Arc6 protein orthologue 2.

KW prokaryotic type; plastid division; Ftn2; Arc6; Arc5; Fzo; plant cell;

KM agronomic; horticultural; crop plant; ornamental plant; woody plant;

XX herbicide target.

OS Arabidopsis thaliana.

XX WO2004001003-A2.

PD 31-DEC-2003.

XX 20-JUN-2003; 2003MO-US019536.

XX 20-JUN-2002; 2002US-0390140P.

XX 09-AUG-2002; 2002US-0402242P.

XX 20-JUN-2003; 2003US-00600070.

XX (UNMS) UNIV MICHIGAN STATE.

XX Oosteryoung KW, Vittha S, Koksharova OA, Gao H;

XX WPI: 2004-082486/08.

XX N-PSDB; ADJ38208.

XX New isolated Ftn2, Arc5 and/or Fzo-like nucleic acid sequences, useful

XX for further characterizing plastid division in plant cells, and in

XX varying agronomic and horticultural characteristics of economically

XX important plants.

XX Disclosure; Fig 8; 287bp; English.

XX This invention relates to novel prokaryotic type or plastid division

XX related genes and proteins. In particular, the invention relates to novel

XX Ftn2 (Arc6), Arc5 and Fzo-like genes and polypeptides. The methods and

XX compositions of the present invention are useful for further

XX characterizing plastid division in plant cells, in order to vary

XX agronomic and horticultural characteristics of economically important

XX plants, such as crop, ornamental and woody plants. They can also be used

XX as herbicide targets. The present sequence is that of a protein which is

XX related to the invention.

XX Sequence 801 AA;

XX Query Match 99.7%; Score 4051; DB 8; Length 801;

XX Best Local Similarity 99.8%; Pred. No. 0;

XX Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 MEALSHVIGIGLSPFQOLCRLPATTKLRSHNTSTTICSAKMDRLSDPNFTSDSSSS 60
DB 1 MEALSHVIGIGLSPFQOLCRLPATTKLRSHNTSTTICSAKMDRLSDPNFTSDSSSS 60
QY 61 PATATTATVLSLPPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFG 120
DB 61 PATATTATVLSLPPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDDALISRQIIOACETLSNPRSREYNEGILLDDEATVITDVPMDKVPALCVLOEG 180
DB 121 FSDDALISRQIIOACETLSNPRSREYNEGILLDDEATVITDVPMDKVPALCVLOEG 180
QY 181 GETEIVLRVGEALLKERLPKSFQODVVLVMAALFLDVSRDAMALDPPDFTGYEFVEAL 240
DB 181 GETEIVLRVGEALLKERLPKSFQODVVLVMAALFLDVSRDAMALDPPDFTGYEFVEAL 240
QY 241 KLIQEGASSLAPDLRAQIDETLEITPRVYLLGLPLGDDVAAKRLNGLSGVRNLTWS 300
DB 241 KLIQEGASSLAPDLRAQIDETLEITPRVYLLGLPLGDDVAAKRLNGLSGVRNLTWS 300
QY 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVEVALVAQA 360
DB 301 VGGGASALVGGITREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVEVALVAQA 360
QY 361 FIGKPHLLQDADKQFQOLQQAQKVMAMEIPAMLYDTRNWEIDGLERGLCALLIGKVD 420

```

```

Db      361 FIGKKPHLLQDADKQFQOLQOAKVAMEIPAMLYDTRNNWEIDPGERGICALLIGKVD 420
Qy      421 CRMWLGIDSDSQYRNPAIVEFYLENSNRDNDLPGLCLETLWLAGVPPRPRTDK 480
Db      421 CRMWLGIDSDSQYRNPAIVEFYLENSNRDNDLPGLCLETLWLAGVPPRPRTDK 480
Qy      481 KFKLGDYDDPMTLSYERVEVQGSPLAAATMARI GAHVKASAMQALOKYPPSRYTD 540
Db      481 KFKLGDYDDPMTLSYERVEVQGSPLAAATMARI GAHVKASAMQALOKYPPSRYTD 540
Qy      541 RNSAEPDQVETVFSVDPVGNNGRDGEPGVFI AEAVRPSENFETNDYAIRAGVSESSVD 600
Db      541 RNSAEPDQVETVFSVDPVGNNGRDGEPGVFI AEAVRPSENFETNDYAIRAGVSESSVD 600
Qy      601 ETTVMSVADMLKEASVKIIAAGVAIGLISLFSQKTYLKSSSFORKDMWSMESDVATI 660
Db      601 ETTVMSVADMLKEASVKIIAAGVAIGLISLFSQKTYLKSSSFORKDMWSMESDVATI 660
Qy      661 GSVRADDSSEALPRMDARTAEINIVSKWOKISLAFGPDHRIEMLPEVLDSGMLKIWTDRAA 720
Db      661 GSVRADDSSEALPRMDARTAEINIVSKWOKISLAFGPDHRIEMLPEVLDSGMLKIWTDRAA 720
Qy      721 ETTAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Db      721 ETTAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Qy      781 RYEVFWSKSGMKITTEGSVLAS 801
Db      781 RYEVFWSKSGMKITTEGSVLAS 801

```

RESULT 3

ADJ38211 ID ADJ38211 standard; protein; 801 AA.

AC ADJ38211;

XX 06-MAY-2004 (first entry)

XX DE Placid division-related Arc6 protein orthologue 4.

XX KM prokaryotic type; Placid division; Ftn2; ARC6; ARC5; Fzo; plant cell;
 XX KM agronomic; horticultural; crop plant; ornamental plant; woody plant;
 XX KM herbicide target.

XX Arabidopsis thaliana.

XX PN MO2004001003-A2.

XX PD 31-DEC-2003.

XX PF 20-JUN-2003; 2003MO-US019536.

XX PR 20-JUN-2002; 2002US-0390140P.

XX PR 09-AUG-2002; 2002US-0402242P.

XX PR 20-JUN-2003; 2003US-0060070.

XX PA (UNMS) UNIV MICHIGAN STATE.

XX PI Oosteryoung KM, Vitcha S, Koksharova OA, Gao H;

XX DR WPI; 2004-082486/08.

XX PT New isolated Ftn2, ARC5 and Fzo-like nucleic acid sequences, useful
 XX PT for further characterizing placid division in plant cells, and in
 XX PT varying agronomic and horticultural characteristics of economically
 XX PT important plants.

XX PS Disclosure; Fig 8; 287bp; English.

XX CC This invention relates to novel prokaryotic type or placid division and
 XX CC related genes and proteins. In particular, the invention relates to novel

CC Ftn2 (ARC6), ARC5 and Fzo-like genes and polypeptides. The methods and
 CC compositions of the present invention are useful for further
 CC characterizing placid division in plant cells, in order to vary
 CC agronomic and horticultural characteristics of economically important
 CC plants, such as crop, ornamental and woody plants. They can also be used
 CC as herbicide targets. The present sequence is that of a protein which is
 CC related to the invention.

XX SO Sequence 801 AA;

Query Match 99.7%; Score 4051; DB 8; Length 801;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy      1 MEALSHVIGLSFQOLCRPLPATTKLRSHNTSTTICSAKMDRLSDNFTSDSSSS 60
Db      1 MEALSHVIGLSFQOLCRPLPATTKLRSHNTSTTICSAKMDRLSDNFTSDSSSS 60
Qy      61 FATATTTATLVSPSPSIDRERHVPPIPIIDFYQLGAQTHELTGIRRAFEARVSKPEQFG 120
Db      61 FATATTTATLVSPSPSIDRERHVPPIPIIDFYQLGAQTHELTGIRRAFEARVSKPEQFG 120
Qy      121 FSDDALISRRQIIQAACETLSNPRSREYNEGILLDBEATVITDVPMDKYPGALCVLQEG 180
Db      121 FSDDALISRRQIIQAACETLSNPRSREYNEGILLDBEATVITDVPMDKYPGALCVLQEG 180
Qy      181 GETEIVLRGEALTKERLPKSFQODVVLVMAALFLDVSRAMLDPPDFITGYEFVEAL 240
Db      181 GETEIVLRGEALTKERLPKSFQODVVLVMAALFLDVSRAMLDPPDFITGYEFVEAL 240
Qy      241 KILQEBGASSIAPDLRAQIDETLEITPRVYLELGLPLGDDYAAKRLNGISGRNIIWS 300
Db      241 KILQEBGASSIAPDLRAQIDETLEITPRVYLELGLPLGDDYAAKRLNGISGRNIIWS 300
Qy      301 VGGGASALVVGILTRKFNMEALFKMTAAEQVDLFAVTPENIAPSEFEYEVALLVAQA 360
Db      301 VGGGASALVVGILTRKFNMEALFKMTAAEQVDLFAVTPENIAPSEFEYEVALLVAQA 360
Qy      361 FIGKKPHLLQDADKQFQOLQOAKVAMEIPAMLYDTRNNWEIDPGERGICALLIGKVD 420
Db      361 FIGKKPHLLQDADKQFQOLQOAKVAMEIPAMLYDTRNNWEIDPGERGICALLIGKVD 420
Qy      421 CRMWLGIDSDSQYRNPAIVEFYLENSNRDNDLPGLCLETLWLAGVPPRPRTDK 480
Db      421 CRMWLGIDSDSQYRNPAIVEFYLENSNRDNDLPGLCLETLWLAGVPPRPRTDK 480
Qy      481 KFKLGDYDDPMTLSYERVEVQGSPLAAATMARI GAHVKASAMQALOKYPPSRYTD 540
Db      481 KFKLGDYDDPMTLSYERVEVQGSPLAAATMARI GAHVKASAMQALOKYPPSRYTD 540
Qy      541 RNSAEPDQVETVFSVDPVGNNGRDGEPGVFI AEAVRPSENFETNDYAIRAGVSESSVD 600
Db      541 RNSAEPDQVETVFSVDPVGNNGRDGEPGVFI AEAVRPSENFETNDYAIRAGVSESSVD 600
Qy      601 ETTVMSVADMLKEASVKIIAAGVAIGLISLFSQKTYLKSSSFORKDMWSMESDVATI 660
Db      601 ETTVMSVADMLKEASVKIIAAGVAIGLISLFSQKTYLKSSSFORKDMWSMESDVATI 660
Qy      661 GSVRADDSSEALPRMDARTAEINIVSKWOKISLAFGPDHRIEMLPEVLDSGMLKIWTDRAA 720
Db      661 GSVRADDSSEALPRMDARTAEINIVSKWOKISLAFGPDHRIEMLPEVLDSGMLKIWTDRAA 720
Qy      721 ETTAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Db      721 ETTAQLGLVYDYLTKLSVDSVTVSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
Qy      781 RYEVFWSKSGMKITTEGSVLAS 801
Db      781 RYEVFWSKSGMKITTEGSVLAS 801

```

RESULT 4

ADJ38209

ID ADJ38205 standard; protein; 801 AA.
 XX AC ADJ38205;
 XX 06-MAY-2004 (first entry)
 DT XX
 XX Plastid division-related Arc6 protein orthologue 3.
 DE XX
 XX prokaryotic type; plastid division; Ftn2; ARC6; ARC5; Fzo; plant cell;
 KM agronomic; horticultural; crop plant; ornamental plant; woody plant;
 KW herbicide target.
 XX Arabidopsis thaliana.
 OS
 XX MO2004001003-A2.
 PN
 XX 31-DEC-2003.
 PD
 XX 20-JUN-2003; 2003MO-US019536.
 PF
 XX 20-JUN-2002; 2002US-0390140P.
 XX PR 09-AUG-2002; 2002US-040242P.
 XX PR 20-JUN-2003; 2003US-00600070.
 XX (UNMS) UNIV MICHIGAN STATE.
 PA
 XX Oosteryoung KW, Vitna S, Kokeharova OA, Gao H;
 XX WPI; 2004-082486/08.
 XX N-PSDB; ADJ38210.
 DR
 XX New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
 PT for further characterizing plastid division in plant cells, and in
 PT varying agronomic and horticultural characteristics of economically
 PT important plants.
 XX
 XX Disclosure; Fig 8; 287bp; English.
 PS
 XX
 XX This invention relates to novel prokaryotic type or plastid division and
 CC related genes and proteins. In particular, the invention relates to novel
 CC Ftn2 (ARC6), ARCS and Fzo-like genes and polypeptides. The methods and
 CC compositions of the present invention are useful for further
 CC characterizing plastid division in plant cells, in order to vary
 CC agronomic and horticultural characteristics of economically important
 CC plants, such as crop, ornamental and woody plants. They can also be used
 CC as herbicide targets. The present sequence is that of a protein which is
 CC related to the invention.
 CC
 SQ Sequence 801 AA;
 Query Match 99.7%; Score 4051; DB 8; Length 801;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MEALSHVIGIGLSPFQQLRLPATTKLRSHNTSTTTCSAGMADRLSDNFNTSSSSSS 60
 DB 1 MEALSHVIGIGLSPFQQLRLPATTKLRSHNTSTTTCSAGMADRLSDNFNTSSSSSS 60
 QY 61 FATATTATVSLPSTIDRBRHVPRIPIDFOVAGATHFLTOSIRAFEARVSKRPQFG 120
 DB 61 FATATTATVSLPSTIDRBRHVPRIPIDFOVAGATHFLTOSIRAFEARVSKRPQFG 120
 QY 121 FSDDALISRRQIIQACETLSNPRSREYNEGGLDDEEATVITDVPMDKYFGALCVLQEG 180
 DB 121 FSDDALISRRQIIQACETLSNPRSREYNEGGLDDEEATVITDVPMDKYFGALCVLQEG 180
 QY 181 GETEIVIRVGEALIKERLPKSPKQDVVLVMAALFDVSRDAMALDPDFITGVEFEAL 240
 DB 181 GETEIVIRVGEALIKERLPKSPKQDVVLVMAALFDVSRDAMALDPDFITGVEFEAL 240
 QY 241 KLOEBGASLAPDLRAQIDETLEETIPRYVLELGLPGDVAARKLNGISGVRLTMS 300
 DB 241 KLOEBGASLAPDLRAQIDETLEETIPRYVLELGLPGDVAARKLNGISGVRLTMS 300

QY 301 VGGGASALVGLTREKFMNEAFLRMTAAEQVDLPVATPSNIPAESFEYEVALLVAQA 360
 DB 301 VGGGASALVGLTREKFMNEAFLRMTAAEQVDLPVATPSNIPAESFEYEVALLVAQA 360
 QY 361 FIKKKPHLLQDADKQOQLOQAKVMAMETPAMLYDRNNMETDPGLERGLCALLIGVDE 420
 DB 361 FIKKKPHLLQDADKQOQLOQAKVMAMETPAMLYDRNNMETDPGLERGLCALLIGVDE 420
 QY 421 CRMWLGIDSDSQRYRPAIVEFLIENSNRDNDLPGCLKLTETWLAGVFPFRFDTKDK 480
 DB 421 CRMWLGIDSDSQRYRPAIVEFLIENSNRDNDLPGCLKLTETWLAGVFPFRFDTKDK 480
 QY 481 KFKLGDYDDPMVLSTLERVVGSSPLAAATMARI GAHVYKASAMQALQKVPESRYTD 540
 DB 481 KFKLGDYDDPMVLSTLERVVGSSPLAAATMARI GAHVYKASAMQALQKVPESRYTD 540
 QY 541 RNSAEPDQVETVPSVDPVNNNGRGEPCGVFAEAVRPSENETNDYAIRAIVSSSD 600
 DB 541 RNSAEPDQVETVPSVDPVNNNGRGEPCGVFAEAVRPSENETNDYAIRAIVSSSD 600
 QY 601 ETTVMSVADMLKEASVKILAAAGVAILGLISLFSQKTFKSSSFORKDMVSSMESDVATI 660
 DB 601 ETTVMSVADMLKEASVKILAAAGVAILGLISLFSQKTFKSSSFORKDMVSSMESDVATI 660
 QY 661 GSVRADSEALPMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEYLDGRMLKIWTDRAA 720
 DB 661 GSVRADSEALPMDARTAEINIVSKWKIKSLAFGPDHRIEMLPEYLDGRMLKIWTDRAA 720
 QY 721 ETKQLGLVYDTYLLKLSVSVSVSADGTRALVATLEESACSLDVHPENNATDVRYTTT 780
 DB 721 ETKQLGLVYDTYLLKLSVSVSVSADGTRALVATLEESACSLDVHPENNATDVRYTTT 780
 QY 781 RYEVFMSKSGMKITGGSVLAS 801
 DB 781 RYEVFMSKSGMKITGGSVLAS 801
 RESULT 5
 ADJ38205
 ID ADJ38205 standard; protein; 760 AA.
 XX
 XX ADJ38205;
 XX 06-MAY-2004 (first entry)
 DT XX
 XX Plastid division-related Arc6 protein orthologue 1.
 DE XX
 XX prokaryotic type; plastid division; Ftn2; ARC6; ARCS; Fzo; plant cell;
 KM agronomic; horticultural; crop plant; ornamental plant; woody plant;
 KW herbicide target.
 XX
 OS Oryza sativa.
 OS
 XX
 XX MO2004001003-A2.
 PN
 XX 31-DEC-2003.
 PD
 XX 20-JUN-2003; 2003MO-US019536.
 PF
 XX 20-JUN-2002; 2002US-0390140P.
 XX PR 09-AUG-2002; 2002US-040242P.
 XX PR 20-JUN-2003; 2003US-00600070.
 XX (UNMS) UNIV MICHIGAN STATE.
 PA
 XX Oosteryoung KW, Vitna S, Kokeharova OA, Gao H;
 XX WPI; 2004-082486/08.
 XX N-PSDB; ADJ38206.
 DR
 XX New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
 PT for further characterizing plastid division in plant cells, and in

PT varying agronomic and horticultural characteristics of economically
 PT important plants.
 XX
 PS Disclosure; Fig 8; 287pp; English.
 CC This invention relates to novel prokaryotic type or plasmid division and
 CC related genes and proteins. In particular, the invention relates to novel
 CC Ftn2 (ARC6), ARCS and Fzo-like genes and polypeptides. The methods and
 CC compositions of the present invention are useful for further
 CC characterising plasmid division in plant cells, in order to vary
 CC agronomic and horticultural characteristics of economically important
 CC plants, such as crop, ornamental and woody plants. They can also be used
 CC as herbicide targets. The present sequence is that of a protein which is
 CC related to the invention.
 XX
 SQ Sequence 760 AA;
 Query Match 43.7%; Score 1775.5; DB 8; Length 760;
 Best Local Similarity 47.9%; Pred. No. 1.9e-150;
 Matches 390; Conservative 119; Mismatches 213; Indels 93; Gaps 16;
 QY 12 SPFOCLPPTATTKRSHTSTTTC-SASKMDRLSDNF-----TSOSSSSPATAT 65
 DB 14 APFAFSLPRPRRRPRPPHPSAACRAASRMARLFPADFHILPTAAPSDPPSPAPAA 73
 QY 66 TTATLVSLPSPIDRPERHVPIDIFYOVLGAQTHFLTDGIRAFARVSKPPQFSDA 125
 DB 74 APSASPPVPLPPDAENSLPQVDPIYKVLGAPPHLGDGIRAFARVAKPPQGYSTDA 133
 QY 126 LISRQILQACETLSNPSRREYNEGLDDEBATEVITDVPMDKVPALCVLOEGGETEI 185
 DB 134 LVGRKQLQIAHDTLMNONSRTQYDRALSENREELTMDIMDK-----EAGEMLA 184
 QY 186 VLKVGALLKERLPEKSKQDYLVMALAFDVSDAMALDPPDITTYEPEYERLAKLQIE 245
 DB 185 VLVTGEQLLDRPKRFKQDVLAMALAYDLISRDAMASPPDVIQCCEVIERLAKLQIE 244
 QY 246 EGASSLAPDLRAOIDEITTEITPRYVELGLPIGDGDAARLMLGSGVRNITLVSGGG 305
 DB 245 DGASNLAPDLISQIDEITTEITPRYVELGLPIGDGDAARLMLGSGVRNITLVSGGG 304
 QY 306 ASALVGLTEKEFNNKAEFLRMTAAEOYDLFVATPSNIPASPEYEVALLVAQAFLGKK 365
 DB 305 IATVGGGFSRRAFNKAEFLRMTAAEOYDLFVATPSNIPASPEYEVALLVAQAFLGKK 364
 QY 366 PHLIQDDAKFOQLQAKWMAEIPAMLYTRNNWEIDFGLERGLCALLGVDECRMTL 425
 DB 365 POFIMMADDLFEOLQKFNIGS---HYAYDN---EMDLALERAFCSLVGVDSKCRMTL 416
 QY 426 GLDSEDSQYRNPAIVEFVLENSN-RDNDNDLPGICXLEETLACVAPRPRDTRDKKPKL 484
 DB 417 GIDNESSPRPKLFLVTNSSLSEENDLPGICXLEETLACVAPRPRDTRDKKPKL 476
 QY 485 GDYDDPMVLSYLERVEVYQGSPLAAATWARIAG-----HYKASMQALQKVP- SR 537
 DB 477 GDYDDPMVLSYLERVEVYQGSPLAAATWARIAG-----HYKASMQALQKVP- SR 536
 QY 538 YTDNRSAEPKQVQRTVSVDPVGNVGRDGEPPVIAEAVRPSSENFETNDYAIRAGVSES 597
 DB 537 QLDPSAMENT-----KDG-PGGYL-----BNFQO----- 559
 QY 598 SVDTTTEMASVADMLKASVYKILAGVALIGLISFQOKYFLKSSSSPORKDMVSSMSD- 656
 DB 560 -----ENNAFHSRNNALKILISAGALFALAAVYAGRT-----LPRKGLPILIRSEH 605
 QY 657 --VATIGSVRADDBEAL-----PRMDARTAEINIVSKQIKSLAFGPHRIEMLEDEVL 707
 DB 606 GSVAVASVSTDDPALDEDPVHLPRMDAKLABDIVRMWSIKSKALGEPHSVASLQEVL 665
 QY 708 DGRMLKWTBAAFTAQIGLVYDYTLKLSVDVTSVADGTRALVEATLEBSACLSDLVH 767
 DB 666 DGNMLKWTBAAFTAEIRHGMFWMEYTLSDVTIDSTIISIDGRRATVEATIDEAGQLTDTVE 725

QY 768 PENNATDVRTYTRYEYFWSK-SGKMTIGSVLAS 801
 DB 726 PRNNDSDYTKTRYEMAFSGKMTIGSVLAS 760
 RESULT 6
 ADJ38203
 ID ADJ38203 standard; protein; 324 AA.
 XX
 AC ADJ38203;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Arabidopsis thaliana AtFtn2 mutated protein.
 XX
 KM prokaryotic type; plasmid division; Ftn2; ARCS; ARCS; Fzo; plant cell;
 KM agronomic; horticultural; crop plant; ornamental plant; woody plant;
 KM herbicide target; mutant; mutain.
 XX
 OS Arabidopsis thaliana.
 XX
 PN W02004001003-A2.
 PD 31-DEC-2003.
 XX
 PF 20-JUN-2003; 2003WO-US019536.
 PR 20-JUN-2002; 2002US-0390140P.
 PR 09-AUG-2002; 2002US-0402242P.
 PR 20-JUN-2003; 2003US-0060070.
 PA (UNMS) UNIV MICHIGAN STATE.
 XX
 PI Oosteryoung KM, Vitsha S, Kosharova OA, Gao H;
 XX
 DR WPI; 2004-082486/08.
 DR N-PSDB; ADJ38135, ADJ38136.
 XX
 PT New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
 PT for further characterizing plasmid division in plant cells, and in
 PT varying agronomic and horticultural characteristics of economically
 PT important plants.
 XX
 PS Disclosure; Fig 2; 287pp; English.
 CC This invention relates to novel prokaryotic type or plasmid division and
 CC related genes and proteins. In particular, the invention relates to novel
 CC Ftn2 (ARC6), ARCS and Fzo-like genes and polypeptides. The methods and
 CC compositions of the present invention are useful for further
 CC characterising plasmid division in plant cells, in order to vary
 CC agronomic and horticultural characteristics of economically important
 CC plants, such as crop, ornamental and woody plants. They can also be used
 CC as herbicide targets. The present sequence is that of a protein which is
 CC related to the invention.
 XX
 SQ Sequence 324 AA;
 Query Match 40.5%; Score 1646; DB 8; Length 324;
 Best Local Similarity 100.0%; Pred. No. 2.2e-139;
 Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MEALSHVIGISLSPFOCLPPTATTKRSHTSTTTCASAKMDRLSDNFSSSSSS 60
 DB 1 MEALSHVIGISLSPFOCLPPTATTKRSHTSTTTCASAKMDRLSDNFSSSSSS 60
 QY 61 FATATTATVSLPSPIDRPERHVPIDIFYOVLGAQTHFLTDGIRAFARVSKPPQFG 120
 DB 61 FATATTATVSLPSPIDRPERHVPIDIFYOVLGAQTHFLTDGIRAFARVSKPPQFG 120
 QY 121 FSDDALISRQILQACETLSNPSRREYNEGLDDEBATEVITDVPMDKVPALCVLOEG 180
 DB 121 FSDDALISRQILQACETLSNPSRREYNEGLDDEBATEVITDVPMDKVPALCVLOEG 180

QY 181 GETEIVLRVGBALLKERLPKSFKODVVLVMAALFLVSRDAMALDPDFITGYFVEEAL 240
DB 181 GETEIVLRVGBALLKERLPKSFKODVVLVMAALFLVSRDAMALDPDFITGYFVEEAL 240
QY 241 KLOEBEASSLAPLRAIODTLEETPRRYVLLELGLPLGDDYAAKUNLNGSGVRNLTMS 300
DB 241 KLOEBEASSLAPLRAIODTLEETPRRYVLLELGLPLGDDYAAKUNLNGSGVRNLTMS 300
QY 301 VGGGASALVGLTRKPMNEAFL 324
DB 301 VGGGASALVGLTRKPMNEAFL 324

RESULT 7
ADJ38247
ID ADJ38247 standard; protein; 768 AA.

AC ADJ38247;
XX
DT 06-MAY-2004 (first entry)

DE Plastid division-related Arc6 orthologue protein 35.

XX prokaryotic type; plastid division; Ftn2; ARC6; ARC5; Fzo; plant cell;
KM agronomic; horticultural; crop plant; ornamental plant; woody plant;
XX herbicide target.

OS Unidentified.

XX W02004001003-A2.

PN 31-DEC-2003.

PD 20-JUN-2003; 2003WO-US019536.

PF 20-JUN-2002; 2002US-0390140P.

PR 09-AUG-2002; 2002US-040242P.

PR 20-JUN-2003; 2003US-0060070.

XX (UNMS) UNIV MICHIGAN STATE.

XX Oeteryoung KM, Vicha S, Kokeharova OA, Gao H;

PI WPI; 2004-082486/08.

DR N-PSDB; ADJ38246.

XX New isolated Ftn2, ARC5 and/or Fzo-like nucleic acid sequences, useful

PT for further characterizing plastid division in plant cells, and in

PT varying agronomic and horticultural characteristics of economically

PT important plants.

XX Disclosure; Fig 8; 287bp; English.

PS This invention relates to novel prokaryotic type or plastid division and

CC related genes and proteins. In particular, the invention relates to novel

CC Ftn2 (ARC6), ARC5 and Fzo-like genes and polypeptides. The methods and

CC compositions of the present invention are useful for further

CC characterizing plastid division in plant cells, in order to vary

CC agronomic and horticultural characteristics of economically important

CC plants, such as crop, ornamental and woody plants. They can also be used

CC as herbicide targets. The present sequence is that of a protein which is

CC related to the invention.

XX Sequence 768 AA;

XX

XX

XX

XX

XX

Query Match 12.9%; Score 524.5; DB 8; Length 768;
Best Local Similarity 24.9%; Pred. No. 1.6e-37;
Matches 212; Conservative 128; Mismatches 286; Indels 227; Gaps 28;
QY 84 VPIPIDEYOVVLAQTHFLTDGIRAFARVSKPPPOGSDALISRRQILQACETLSNP 143
DB 1 VRIPUDYRILGLPLAASSEGLRQAYSDRIVQLRRREISQALISSKRLIEATVVLSDP 60

QY 144 RSRRREINGL-----DDEEATVIT-----DVPMDKVRGALCVLQ 178
DB 61 KORSTYDQLYLAAYDPDNLAAAVAOENRTESTKSGSDTQSLGIEITDDELGALLILQ 120
QY 179 EGETEIVLRVGBALL--KERLPKSFQ-----DVVLVMAALFLVSRD- 220
DB 121 ELGETEIVLRVGBALL--KERLPKSFQ-----DVVLVMAALFLVSRD- 220
QY 221 -----AMALDPDFITGYFVEEALKLOEBEASSLAPLRAIODTLEETPRRYV 271
DB 181 MOGHYENNAISLE-----TGQ-----LVREG-----LFSSIQARIQADLYKLRERYI 225
QY 272 LELLGLPLGDDYAAKUNLNGSGVRNLTMSVG--CGASALVGLTR--KPMNEAFLRMT 327
DB 226 LELLGLPLGDDYAAKUNLNGSGVRNLTMSVG--CGASALVGLTR--KPMNEAFLRMT 327
QY 328 AAEQVDFVATPSPNIPAESFEYVEVALVAQAFIGKKPHLLDDADQFOQLQAKMMAM 387
DB 284 VAEOHKLFEAQSRR--SSAVATYLAAYALTARGPAQOPALIRQAKOMLVRLGRQ--- 337
QY 388 EIPAMLYDTRNNWEIDFGLEGLCALLIGKVDCEKRMVLGDSQYRNPAIVEFVLENS 447
DB 338 -----DVHLEQSLCALLIGQTEATRYLELSQ---YE---ALAFIREKS 376
QY 448 NRDDNDLPELCKLETWLAGVVPFRPRDYDKKFKLGDYDDPMVLSYLERVEVVOGSP 507
DB 377 -QDSPDLLPELCLYAEQWLOHEVFPFRDLANOQAFLKDYFANQOVQAYILE----- 426
QY 508 LAAATWARIAGAHVYKASAMQALOKVPPRYTBRNSAEPR-----DVOETVFSV 556
DB 427 --ALPTDAQTTEWAVYVNPQYFPQAKAKNTHFNHNSKTTSASRHSRVPRDLPET--- 480
QY 557 DPVGNVNGRDEBGPVIAEAVRPSSENEETNDYAIRAGVSSSVDETTVENSVAQMLK-- 613
DB 481 -----PTK--ETSEYVNFSPMMWSSSGSISKEVPALBRMSRGT 516
QY 614 ----EASVKTLAAG----- 623
DB 517 NOHLNGSAKSAASGHQKRRRRKPTSPASRERI PDNRPHSRPRRRRTFANTTEGKTRLV 576
QY 624 --VAIGLISL-----FSQKFLKSSSPORKDMVSMESDVATIGSVRADSE 669
DB 577 WRVFIISVLSILVWVLAATTTFGWLKNLFFPQBPRLQLFVQINQRPPLRPDPNRKRESE 636
QY 670 ALPRMDARTENIVSKQKIKSLAFGPDHRIEMLPVLDGRMLKIWTDRAETAGLGLVY 729
DB 637 EGPLTNAE-AEEYIHTWLSTKKAALGPNHINNLEQILTSALSQWR-LIAQNKLDNRY 694
QY 730 ---DYTLKLKLSVDSVVSADGTALVEATLEBACLSDLVHPENMATDVATTTRYEVFW 786
DB 695 RKFDHSLIKIESVERKIGLFAD--RAAVEATVKEVTOLYENNOFQNSND--KLRRYDILR 750
QY 787 SKSGWKTEGGSVL 799
DB 751 ERGKWRIGSTSV 763

RESULT 8
ADJ38274

ID ADJ38274 standard; protein; 789 AA.

AC ADJ38274;

XX 06-MAY-2004 (first entry)

DT 06-MAY-2004 (first entry)

DE Plastid division-related Arc6 orthologue protein 58.

XX prokaryotic type; plastid division; Ftn2; ARC6; ARC5; Fzo; plant cell;

XX agronomic; horticultural; crop plant; ornamental plant; woody plant;

XX herbicide target.

OS Trichodesmium erythraeum.

XX

PN WO2004001003-A2.
 XX 31-DEC-2003.
 PD 20-JUN-2003; 2003WO-US019536.
 XX 20-JUN-2002; 2002US-0390140P.
 XX PR 09-AUG-2002; 2002US-0402242P.
 PR 20-JUN-2003; 2003US-00600070.
 XX (UNMS) UNIV MICHIGAN STATE.
 XX Oosteryoung KM, Vitha S, Koksharova OA, Gao H;
 PI WPI: 2004-082486/08.
 XX N-PSDB; ADJ38273.
 DR
 XX New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
 PT for further characterizing plastid division in plant cells, and in
 PT varying agronomic and horticultural characteristics of economically
 PT important plants.
 XX
 PS Disclosure; Fig 8; 287pp; English.
 CC This invention relates to novel prokaryotic type or plastid division and
 CC related genes and proteins. In particular, the invention relates to novel
 CC Ftn2 (ARC6), ARCS and Fzo-like genes and polypeptides. The methods and
 CC compositions of the present invention are useful for further
 CC characterizing plastid division in plant cells, in order to vary
 CC agronomic and horticultural characteristics of economically important
 CC plants, such as crop, ornamental and woody plants. They can also be used
 CC as herbicide targets. The present sequence is that of a protein which is
 CC related to the invention.
 CC
 XX Sequence 789 AA:

Query Match 12.3%; Score 498.5; DB 8; Length 789;
 Best Local Similarity 23.7%; Pred. No. 3.8e-35;
 Matches 203; Conservative 132; Mismatches 306; Indels 215; Gaps 31;

QY 84 VPIIDYVYVLAQTHLTDGIRAFARVSKPPQFGFSDALISRQIILQACETLSNP 143
 DB 1 VRILDYRIRIGLPIQATAEQLRQAHQDRTOQPPRRESEATYVARQGLIDEAVAVLCDF 60
 QY 144 RSREYNEGLI----- 154
 DB 61 EQRQTYDGNFLAKTYEPIVEELNPSQINPQAOKEKTELTRETVLEIASKOLKKRTS 120
 QY 155 -----DDEATVITDVPMKVPKALCVLQGGGETEIVLVGEA-LIKERL----- 198
 DB 121 YQNETAASDPHSNTPSIEIYEPQVGAIIILHLEGYELVLTHTPYLQNNSTITKDG 180
 QY 199 -----PKSFKQDVVLVMAALAFDVSADMAALDPDFITG-YEFVEALK-----LLOEGA 248
 DB 181 RFGDPALVLPVVLTVALNLELGR-----EMQGYSAFTALEAGLILRE----- 230
 QY 249 SSIALPDLPAQIDETLEETPRYVLELLGLPGDDYAAARLNGLSGVNIIIVSG3-GGAS 307
 DB 231 -NLFTVQIRGEIQADLYKLRPYRIMELALP--BEIALDRSGLEILQMLNERGGIDGOG 287
 QY 308 ALVAGGLTRE---KMNNEAFLEMTAAEQVDLEVA---TPSNIPASFYEVYVALVLAQAF 361
 DB 288 EDSGGIGLEDFLKVQOQLROYLTTRAEQKLEAEALRSVAGA---YLAVTPLAQGF 342
 QY 362 IGKPEHLLDQADKFOQLQAKVWAMEIPAMLYDTRNMEIDFGLERGLCALLIGKVDCE 421
 DB 343 AQKQPAPIRKAKMLMQLGRSG-----DVNLKSVCLALLIGQTEBA 383
 QY 422 RMLGLDSEDSQYRNPAIVLEVLENSRNDNDLPGCLKLETLWLAGVVPFRPDYDKK 481
 DB 384 SRSLLESHENP-----LSFIKENSQOSP-DLPLGLCLVAEHMLTEBEVPHFPLDLSKS 436
 QY 482 FKLDYDDDDPVNLSYLERVEVVGSPPLAATAATMARIGAEHNKASMAQLQVPPSRRTYDR 541

DB 437 ASLDYFADQGVQAYLE-----ALPTEAEVANNQVWVQPRR-----SNHKK 478
 QY 542 NSAPKXVQ-----ETVFSVDPVGNV---GRGEPGVFIAEAVRPENFETDYA 589
 DB 479 QMFDPKLEKLVNSDLEDKDISRYDATATGIVASGQSSNLLGASSDGLQLEKSS-S 537
 QY 590 IRAG-----VSESSV-----DETVEEM-SVADMLKEASVK----- 618
 DB 538 TRGSPKQVTTKSSHYLGKIREKISGLPERNESTISGGLPQSIGHSRRTSARRE 597
 QY 619 -----ILAGVA---IGLISLFSQKY--FLKSSSSPQRKDMVSSMSDVATIGSVRAD 667
 DB 598 VKFGRILILIAIVGFLIGLITKITIGMLVNLGWEREKLMQLDRPPIEIPDRVN 657
 QY 668 SEALPRMDARTAEIVSKWKIKSLAGPDHRIEMLEVDGRMLKTV--IDRAAEATQL 725
 DB 658 LAAGPITKEVARRTTISWMDIKASALGPNNKIEQLPNITVPEALSRPLPANNLKOEKS 717
 QY 726 GLVYDYTL---LKLSDVSVTVSADGTRALVEA-TLESACTSD---LVHPENNATDVRT 777
 DB 718 YRREYHDLAISNIMMS-----NTNSMLAQVDKAVIEKVEYSUNGRLTNTNMENLPVRY 771
 QY 778 YTRYEVFWMSKSGWKI 793
 DB 772 DLVVKSGKMQISNMKV 787

RESULT 9
 ADJ38244
 ID ADJ38244 standard; protein; 798 AA.
 XX AC ADJ38244;
 XX DT 06-MAY-2004 (first entry)
 XX DE Plastid division-related Arc6 orthologue protein 33.
 XX KW prokaryotic type; plastid division; Ftn2; ARCS; ARCS; Fzo; plant cell;
 KW agronomic; horticultural; crop plant; ornamental plant; woody plant;
 KW herbicide target.
 XX OS Anabaena sp.
 XX PN WO2004001003-A2.
 XX PD 31-DEC-2003.
 XX PE 20-JUN-2003; 2003WO-US019536.
 XX PR 20-JUN-2002; 2002US-0390140P.
 PR 09-AUG-2002; 2002US-0402242P.
 PR 20-JUN-2003; 2003US-00600070.
 PA (UNMS) UNIV MICHIGAN STATE.
 XX Oosteryoung KM, Vitha S, Koksharova OA, Gao H;
 PI WPI: 2004-082486/08.
 XX
 PT New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
 PT for further characterizing plastid division in plant cells, and in
 PT varying agronomic and horticultural characteristics of economically
 PT important plants.
 XX
 PS Disclosure; Fig 8; 287pp; English.
 CC This invention relates to novel prokaryotic type or plastid division and
 CC related genes and proteins. In particular, the invention relates to novel
 CC Ftn2 (ARC6), ARCS and Fzo-like genes and polypeptides. The methods and
 CC compositions of the present invention are useful for further
 CC characterizing plastid division in plant cells, in order to vary
 CC agronomic and horticultural characteristics of economically important

CC plants, such as crop, ornamental and woody plants. They can also be used
 CC as herbicide targets. The present sequence is that of a protein which is
 CC related to the invention.
 XX

SO Sequence 798 AA;

Query Match 12.2%; Score 497; DB 8; Length 798;

Best Local Similarity 24.6%; Pred. No. 5.3e-35;
 Matches 207; Conservative 139; Mismatches 301; Indels 196; Gaps 33;

```

QY      84 VPIPIDFYOVUGAOTHLTDGIRAFARVSKPPQGFSDALISRQIIQAACETLSNP 143
      111 |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
      11 VRIPLDYRILGLPLASADQLRQAYSDRIVQLRRRYSQALASRKOLIEAVVLSDP 70
      144 RSREY-----NEGLDDEATVITDVPMDKVPGLCV 176
      71 KERSSTYQVLYLAHAYPDNNAATTVAVENRGDSNNGHFDVQSLSI--EVSSEELIGALLI 128
      QY      177 LQEGGETEIVLRVGEALL-----KERLPKSF-----KQDVVLVMAIAFLDVS 219
      129 LQELGEYELVLYKGRNYLGNQNGTASTRNGNHRTPPEFLDSSRPDILLTVALLASLELGR 188
      QY      220 D-----AMALDPDPFITGYEFVEEALKLQEGGASSLAPDLRAQIDETLEETPR 269
      189 EQWQGHYENAAALSLR-----TGQ-----VLFSEG--IPSVQAEIQADLYKLRPY 233
      QY      270 VYLELGLPLGDDVYAAKRLNGLSGVNIIILMSVGG--GASALVGLTRE--KFMNEAFLR 325
      234 RIEELALP--QEKTERHQGLDLSQSLIDRGSIDGTGDSGLNIDDLRFIQQLRHH 291
      QY      326 MTAAEQVDFVATPSPNI PASSEFEYVALALVAQAFIGKKPHLLQDADKOPQLOQAKVM 385
      292 LTVAEQKHLFDG--BSKRP--AVATYLAIVASIRAGFORQALIRHAKQILMRLSKQ-- 347
      QY      386 AMELPAMLVTRNNWEIDFGLEBGLCALLGKVDCEMMGLDSEBDQYRNPAAVEVLE 445
      348 -----DVHLEQSLCALLGQTEAEATRVLELSQ--YENALALI-----R 383
      QY      446 NSNRDNDNDLPGCLKLETWLAGVPRPRTXOKKFKLGDYDDPMWLSYLERV----- 500
      384 EKSQDSFDLPLGLCLYAEQWLONEVFPHFDLSRQOASLKDYIFANQOVQAYLEMLPDAE 443
      QY      501 -----EVVQSSPLAAATMAR--IGAEL--VYASAMQALQKVPSP--RYTD 540
      444 TTMEWAVINRQSPSQPRGNSVSGTPVAKRPVGRANRGEASTRPVQRSHSPSEVNRQFH 503
      QY      541 RNSAEPKVOGETVPSVDVPGNNVGRDGE-----PGVFIABAVRPSSENFETND 587
      504 QNRTTPELEPETSNNRRRPESSNFTTARENTSTDTAYTDNYPPEIIPVERASRP----- 555
      QY      588 YAIRAGVS---ESSVDETV-----EMSVADMUKKASVKILAA 622
      556 --VQPGVSGTQSTPPRQTPKRRRRKRPQAVVNRGSHIHQORQSPSTLGRKTRILMIVL 613
      QY      623 GVALIGLSLP---SQRY-FLKS-----SSSFQRKDWVSSMESDVATIGSVRADSEALPR 673
      614 G-SLGGILFMLIVSTTFEWMKQNVFPAPASLQGEQLSIQISQPLEIIPDKNAQIQSPSEVS 672
      QY      674 MDARTANIVSKWOKISLAFGPDRHTEMPEVLDGMLKIWTDRAAETQOLGLVITYTL 733
      673 LTEETIARKIIEMLMATASALGAHEKIESINEILITGSLQWMLIALQADNRHRRPS- 731
      QY      734 LKISVSVTVS--ADGRTALVEATLESAGLSDLVH--PENNAATDVRTYTRTYEYFWMKSG 790
      732 HSYKVDISIKSDIDPNRPAVGATVR--LTQFYENGQKGSDEK--LRYVELIIRODDI 787
      QY      791 WKI 793
      788 WRI 790
  
```

RESULT 10
 ADJ38245

```

ID      ADJ38245 standard; protein; 798 AA.
XX
XX      AC      ADJ38245;
XX
XX      DT      06-MAY-2004 (first entry)
XX
XX      DE      Plastid division-related Arc6 orthologue protein 34.
XX
XX      KW      prokaryotic type; plastid division; Ftn2; ARC6; ARC5; Fzo; plant cell;
      KM      agronomic; horticultural; crop plant; ornamental plant; woody plant;
      KW      herbicide target.
XX
XX      OS      Anabaena sp.
XX
XX      PN      WO2004001003-A2.
XX
XX      PD      31-DEC-2003.
XX
XX      PF      20-JUN-2003; 2003WO-US019536.
XX
XX      PR      20-JUN-2002; 2002US-0390140P.
      PR      09-AUG-2002; 2002US-0402242P.
      PR      20-JUN-2003; 2003US-00600070.
XX
XX      PA      (UNMS ) UNIV MICHIGAN STATE.
XX
XX      PI      Oosteryoung KW, Vitsha S, Koksharova OA, Gao H;
      XX      WPI; 2004-082486/08.
XX
XX      PT      New isolated Ftn2, ARC5 and/or Fzo-like nucleic acid sequences, useful
      PT      for further characterizing plastid division in plant cells, and in
      PT      varying agronomic and horticultural characteristics of economically
      PT      important plants.
XX
XX      PS      Disclosure; Fig 8; 287bp; English.
XX
XX      CC      This invention relates to novel prokaryotic type or plastid division and
      CC      related genes and proteins. In particular, the invention relates to novel
      CC      Ftn2 (ARC6), ARC5 and Fzo-like genes and polypeptides. The methods and
      CC      compositions of the present invention are useful for further
      CC      characterizing plastid division in plant cells, in order to vary
      CC      agronomic and horticultural characteristics of economically important
      CC      plants, such as crop, ornamental and woody plants. They can also be used
      CC      as herbicide targets. The present sequence is that of a protein which is
      CC      related to the invention.
XX
XX      SQ      Sequence 798 AA;
XX
XX      Query Match 12.2%; Score 497; DB 8; Length 798;
      XX      Best Local Similarity 24.6%; Pred. No. 5.3e-35;
      XX      Matches 207; Conservative 139; Mismatches 301; Indels 196; Gaps 33;
  
```

Qy	326	MTAAQGVDLFVATTPSNTPAESFEVYVALALVAQATIGKKPHLLQDADKQFQOLQQAkvM	385
Dd	292	LTVAAQHKLPDQ-ESKRPS-AVATYLAVYASIRGFTOROPALIRNAKQIIMKLSKQ--	347
Qy	386	AMEIYAMLVDYTNMNEIDFGLRGICALLIGKDECRMMIIGLSEDSQYENPAIVEVLE	445
Dd	348	-----DVHLEOSLCALLIGQTEAKTRVLEISQE--YALALTI-----R	383
Qy	446	NSNRDNDNDLPGICLKLETWLGAVPEPRPDTKDKFKLGDIYDDPWLSTYLERV----	500
Dd	384	EKSQSPDLLPGLCYIABQWLNQNEVPHFRDLSRQQAISLKDYPANQOQVAYLEALPYDAE	443
Qy	501	-----EYVQSPFLAAATMAR-IGAEH--VKASMAQALQKVPS--RYTD	540
Dd	444	TTNEMAVINROFSQPRGNSYSGTFVAKRPVGKANRPGEASTRPVQSRHPSBKNRQFH	503
Qy	541	RNSAPKQVQETVFSVDVPGNNVGRGE-----PGVFLAEAVRPSNPETND	587
Dd	504	QNRTPDPPELPETSNHRRPSSNFTTARENIISTDAYTDTNYPPEIYVERASRP-----	555
Qy	588	YAIRAASV---ESSYDETV-----EMSVADMKEASVYKILAA	622
Dd	556	--VQGVGSYGTYOSTPPTROTPKRRKRKKQPAVYVNGHSHIQQRPSSTGLGRKRLMIYV	613
Qy	623	GVAIGLISLF---SOKY-FLKS---SSSFQKDMVSSNESDVATIGSVRADSEALPR	673
Dd	614	G-SLGGILLFWLIVSTTFGMLKNKVPFAPSILQEQSLSIQISCPLEIPDKNMQIQSEVPS	672
Qy	674	MDARFRANIVSAKQKIKSLAFPGDRIEMLPYLDGRMLKIYWDRAAEFTQQLGVYDYL	733
Dd	673	LTERFARKIENWLATKASALGAENKIESLINELTGSALSQWPLIALQDRAENRREYS-	731
Qy	724	LKLSTDSVTVS-ADGTRALVEATLEESACTLDLVH--PENNAATDVRYTTRVFWFSKSG	790
Dd	732	HSVKVDSISKSDIDPRKASVGATVR--LTQPYENGQKKSSEDER-LRVYELIINODDI	787
Qy	791	WKI	793
Dd	788	WRI	790
RESULT 11			
ID	ADJ38242		
XX	ADJ38242	standard; protein; 631 AA.	
XX	ADJ38242;		
XX	AC		
XX	DT		
XX	06-MAY-2004	(first entry)	
XX	DE		
XX	Placid	division-related Arc6 orthologue protein 32.	
XX	KW	prokaryotic type; placid division; Fcn2; ARC6; ARCS; Fzo; plant cell;	
XX	KM	agronomic; horticultural; crop plant; ornamental plant; woody plant;	
XX	XX	herbicide target.	
XX	OS		
XX	Synechococcus	sp.	
XX	WO200401003-A2.		
XX	31-DEC-2003.		
XX	20-JUN-2003;	2003WO-US019536.	
XX	20-JUN-2002;	2002US-0390140P.	
XX	09-AUG-2002;	2002US-0402242P.	
XX	20-JUN-2003;	2003US-00600070.	
XX	(UNMS)	UNIV MICHIGAN STATE.	
XX	Osteryoung KM,	Vittha S, Kokesharova OA, Gao H;	
XX	WPI; 2004-082486/08.		
XX	N-PSDB; ADJ38243.		
DR			

xx New isolated Ftn², ARC5 and/or Fzo-like nucleic acid sequences, useful
PT for further characterizing plastid division in plant cells, and in
PT varying agronomic and horticultural characteristics of economically
PT important plants.
xx
xx
PS Disclosure; Fig 8; 267pp; English.
xx
xx This invention relates to novel prokaryotic type or plastid division and
CC related genes and proteins. In particular, the invention relates to novel
CC Ftn² (ARC6), ARC5 and Fzo-like genes and polypeptides. The methods and
CC compositions of the present invention are useful for further
CC characterising plastid division in plant cells, in order to vary
CC agronomic and horticultural characteristics of economically important
CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is that of a protein which is
CC related to the invention.
SQ Sequence 631 AA;
xx

Query Match	10.9%	Score 442;	DB 8;	Length 631;
Best Local Similarity	24.5%	Pred. No. 3,3e-30;		
Matches	189;	Conservative 118;	Mismatches 262;	Indels 204; Gaps 30
QY	IPIDFYVYLAQTHFLTDGIRAFARARSKPKPFGFSDPALISRRQILQACETLSNPRS	145		
DB	3 IPIDYTRILTCGVQASADKLAEHYNDRLNQSSEHSESLAQARQLLEAAIAELSDPQ	62		
QY	146 RREYNE-----GLDDEEAITYTDPWDKVPGALCVLQGGETEYLRVGEALL-----	194		
DB	63 RDRYDRPFQGGILEAIEPFSLELED--MORI-GALLILLETGEYDVVSQIAELELPPDYAS	119		
QY	195 KERLPKSKQDPVLYMALAFIDVSDNDAALDDPDDITGTEFEFBEALLKILQGEA-----	248		
DB	120 ABRVDQFARGIATLALIASQSLGKCRQ-----QGLEYDQAQHFGNS	162		
QY	249 -SSLA-----PDLRAQIDETLEETIPRAYLELGLPLGDDYAKRLNGLSGVNLT---W	299		
DB	163 QSALADHQRPELSTTHQEGQQLRPYHILBELAQLPTAD--SDRQGGILLIQAMLDRO	220		
QY	300 SVGGGASALVGGLTREKFMNEAFLR-----MTAAEQVDLF---VATPSNIPAESFEVTE	351		
DB	221 GIEGGDGG--SGLTLDNPL--MFLQQRNGYITTLAEQQLFESEARPS--PAASF---F	271		
QY	352 VALALVAQAFIGGKKEHLLQDADKQEQQIQAKVMAEIPAMLYDRNMWEIPGFERGC	411		
DB	272 ACTYTLIAGPCDHOSSLIRASLILHEKS-----RMDVHIQDAIA	312		
QY	412 ALLIGKVECEMMVGLDSEED-----SQYRNPAIVEFVLENSRNDNDLPGCLKLETWLA	467		
DB	313 SLLGQPEAEALL--VQSGDEBTLISQIRALQGEAL-----LYGLCRFETWLA	360		
QY	468 GVAFPRFPDITDKFKLGDYDDPENVLSTLERVEVVOGSPLAALAAIMARIAGAHVAKASM	527		
DB	361 TKVFEDFPDLTKERTAPLPQPFYDDPDPVQYTLDAIVEL-----	396		
QY	528 QALQGVFFSRITDRASAPKQVQEVFVSVD-----VGNVGRDGEPGVFLAEAVRPS	580		
DB	397 -----PSDLMPPPLPEBPELVRSLSLAKELPTPATPG-----VAPP	432		
QY	581 ENFETNDYAIRAGVSESSVDETTVMSSVADMLKEASVKYLLAAGVAYIGLISLFSOKYFLKS	640		
DB	433 PRRRRRDSSEKPARAKRKLPLPWIGLV-----VVVLGGGTGV-----WAMRS	475		
QY	641 SSSFORKQVSMESDVAITIGSVRADDSBALP-----RMDARTANIVSKWQKTS	691		
DB	476 RSN-----STPPTPPVVGVTLPBEAVAPAPAPATYVALDRAQAEYVALQNMIAATA	524		
QY	692 LAFGDRHRIEMLPVLDGRMLKIMTDRAEET--OLGLVYDITLLKLSVDSYTVASDGR	749		
DB	525 AALGQYQYRDRLATYLTGSEVLTQWTGQFSSQANTLQTSQFDH---KLTADSYQLSGDQR	581		
QY	750 ALVATVLEESACSLDVHPENNATDVRT---YTTREVEVFWSKGMKITEGSVAL	799		

Db 582 AVQAKVDE----VEQYVRGDQLELTRRDGLVIRYQLVRENNIMKIASISLV 630

RESULT 12
ADJ38132
ID ADJ38132 standard; protein; 631 AA.

XX AC ADJ38132;
XX DT 06-MAY-2004 (first entry)
XX DE Synecchococcus ftn2 protein sequence.
XX KM prokaryotic type; plactid division; Ftn2; ARC6; ARCS; Fzo; plant cell;
XX KM agronomic; horticultural; crop plant; ornamental plant; woody plant;
XX KM herbicide target.
XX OS Synecchococcus sp.
XX PN MO2004001003-A2.
XX PD 31-DEC-2003.
XX PF 20-JUN-2003; 2003MO-US019536.
XX PR 20-JUN-2002; 2002US-0390140P.
XX PR 09-AUG-2002; 2002US-0402242P.
XX PR 20-JUN-2003; 2003US-00600070.
XX (UNMS) UNIV MICHIGAN STATE.
XX PI Oesteryoung KM, Viltha S, Kokeharova OA, Gao H;
XX DR WPI: 2004-082486/08.
XX DR N-PSDB; ADJ38131.

PT New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
PT for further characterizing plactid division in plant cells, and in
PT varying agronomic and horticultural characteristics of economically
PT important plants.

XX Claim 1; SEQ ID NO 5; 287bp; English.

CC This invention relates to novel prokaryotic type or plactid division and
CC related genes and proteins. In particular, the invention relates to novel
CC Ftn2 (ARC6), ARCS and Fzo-like genes and polypeptides. The methods and
CC compositions of the present invention are useful for further
CC characterizing plactid division in plant cells, in order to vary
CC agronomic and horticultural characteristics of economically important
CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is that of a protein which is
CC related to the invention.

XX SQ Sequence 631 AA;

Query Match 10.9%; Score 442; DB 8; Length 631;
Best Local Similarity 24.5%; Pred. No. 3.3e-30;
Matches 189; Conservative 118; Mismatches 262; Indels 204; Gaps 30;

QY 86 IPIDPVOVLGAQTHFLTDGIRRAFEARVSKPQFGSDALISRRQLQAACETLSNPS 145
DB 3 IPIDVYRILCVGQASADKLAEVYRDLNQSFSHFSBLAQARRQLRAIAELSDPEQ 62
QY 146 RREYNE-----GLDDEAVITDVPPWDKYPGALCVQEGGEHIVARVEALL----- 194
DB 63 RDRYDRRPFQGGLEALPSELEED--WQRI-GALLILLEGKYDRVSQLEABELLPDYDAS 119
QY 195 KERLPKSFKODVVLVVMALAFLDVSRDAMALDPDPFITGYEFVEBALTLLOBEGA----- 248
DB 120 AEVRDQPARBDIALATLSQGSISLGRGQ-----GSLYEQAAQHGRS 162
QY 249 -SSLA-----PDLRAQIDETLEETPRYVELLGLPLGDDYAAKRLNGLSGVENIL--W 299

Db 163 QSALAHQRPPELSRTHQEQQLRPYRILERLAQPLTAD--SDROGQLLLQAMLDROQ 220
QY 300 SVGGGASALVGGITREKFMNEBALR-----MTAAEQVDLF---VATPSNIPAESFEVYE 351
DB 221 GIEGPDG--SGITLDNPL--MFLQIRGYTLTAEQQLFESBARPS--PAASF---F 271
QY 352 VALALVAQAFIGKKPHLLQDADKQFOQLQQAQKNAMEIPMLYDTRNNWEIDFGLERGLC 411
DB 272 ACYTLIARGFCDHPSLIHRASILLHELKS-----RMDVHIEQAI 312
QY 412 ALLIGKVDCEKMLGLDSED-----SQYRNPAYIEPVLENSNRDNDLPEGLCTLEETWLA 467
DB 313 SLLIGQPEEAEALL--VQSDDEETLSQIRALAQGEAL-----IVGLCRFTETWLA 360
QY 468 GVPFPRPDKDKKFKLGDYDDPMVLSTLERVEVQGSPLAAATWARIQAEHVKSAM 527
DB 361 TKVPFDPDRDKERTAPLQYFDDPDVQTYIDALVEL----- 396
QY 528 QALQKVPFSRYTDRNSAPKDVQETVFSYDP-----VGNVYGRDGPVFTAAVAVPS 580
DB 397 -----PDLMPITLPVPEPLEVRSSILAKELPTPATPG-----VAPF 432
QY 581 ENFETNDYAIRAGVSESSVDETTENSVDADMLKASVXIIAAGVAIGLISLPSOKYFLKS 640
DB 433 PRRRRRDRSERPARTRPLPFWIGLV-----VVVLGGGTGV-----WAMRS 475
QY 641 SSSFORKDWVSMESDVATIGSVRADDSALP-----RMDARTANIVSKQKIKS 691
DB 476 RSN-----STPPTPPVQVQTLPEAVPSPAPVVALDRAQAEVTLQWMLAAKA 524
QY 692 LAFGPDHRIEMLPBVLIDGRMLKIWTDRAAETA--QLGLVYDYTLKLKSVSVTSADGTR 749
DB 525 AALGPQYDRDLATVLTGVLQWGFSSQQAQNTQLTSGFDH---KLTVDVSQLSDDQK 581
QY 750 ALVEATLEBSACSDLVHPENNAITDVRT---YTRRYEVFWSKSGWKITTESYL 799
DB 582 AVQAKVDE----VEQYVRGDQLELTRRDGLVIRYQLVRENNIMKIASISLV 630

RESULT 13
ADJ38240
ID ADJ38240 standard; protein; 631 AA.

XX AC ADJ38240;
XX DT 06-MAY-2004 (first entry)
XX DE Plactid division-related Arc6 orthologue protein 31.
XX KM prokaryotic type; plactid division; Ftn2; ARC6; ARCS; Fzo; plant cell;
XX KM agronomic; horticultural; crop plant; ornamental plant; woody plant;
XX KM herbicide target.
XX OS Synecchococcus sp.
XX PN MO2004001003-A2.
XX PD 31-DEC-2003.
XX PF 20-JUN-2003; 2003MO-US019536.
XX PR 20-JUN-2002; 2002US-0390140P.
XX PR 09-AUG-2002; 2002US-0402242P.
XX PR 20-JUN-2003; 2003US-00600070.
XX (UNMS) UNIV MICHIGAN STATE.
XX PI Oesteryoung KM, Viltha S, Kokeharova OA, Gao H;
XX DR WPI: 2004-082486/08.
XX DR N-PSDB; ADJ38241.
XX

PT New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful
PT for further characterizing plastid division in plant cells, and in
PT varying agronomic and horticultural characteristics of economically
PT important plants.

XX Disclosure; Fig 8; 287pp; English.

CC This invention relates to novel prokaryotic type or plastid division and
CC related genes and proteins. In particular, the invention relates to novel
CC Ftn2 (ARC6), ARCS and Fzo-like genes and polypeptides. The methods and
CC compositions of the present invention are useful for further
CC characterizing plastid division in plant cells, in order to vary
CC agronomic and horticultural characteristics of economically important
CC plants, such as crop, ornamental and woody plants. They can also be used
CC as herbicide targets. The present sequence is that of a protein which is
CC related to the invention.

XX Sequence 631 AA;

Query Match 10.9%; Score 442; DB 8; Length 631;

Best Local Similarity 24.5%; Pred. No. 3.3e-30;

Matches 189; Conservative 118; Mismatches 262; Indels 204; Gaps 30;

QY 86 IPIDFYVGAQTHTLTDGIRAFARVSKPPGFSDDALISRQIIQAACETLSNPRS 145
DB 3 IPIDYHILCVGQASADKLAEYRDLNQGSPHSFSLAQARQLLEAIAELSDPEQ 62
QY 146 RREYNE-----GLDDEATYITVPMDKVPALCVLOEGGETIYLRVGEAL----- 194
DB 63 RDRYDRRFPQGLAEIPELSLEED--WQRI-GALLILLEIERYVSQLABELLPDYDAS 119
QY 195 KERLPKSPKODVVMALAFDVSADAMALDPPFITGYEVEBALKULOEGA----- 248
DB 120 AEVADQPARGDIALAIALSQSLGECRQ-----QGLYEQAAQHFRS 162
QY 249 -SSIA-----PDLEAQIDETLEITPRVYELGLPLGDDYAAKRLNGSGVRNLT--W 299
DB 163 QSLADHQRPELISRTLHQEGQQLRPYRIERLAQPLRAD--SPRQGLLLQMLMDRQ 220
QY 300 SVGGGASALVGLITREKFNMEALR-----MTAAEQVDF--VATPSNIPASFEYIE 351
DB 221 GIEBPGRDGG--SGTLTNNFL--MFLQIRGYTLAEQQLFESEARPRS--PAASF--F 271
QY 352 VALALVQAFTGKKEPHLLQDADKQFQQAQKWMAMEIPALYTRNNWEIDFGELGCL 411
DB 272 ACYTLIARGFDHPSLIHRASLLIHEKLS-----RNDVHIQMA 312
QY 412 ALIIGKYDECRMWGLDSED-----SQYRNPAIVEFVLENSNRDNDLPGCLLETWLA 467
DB 313 SILIQGEAEAL--VQSDBETISQIRALAQGEAL-----IVGLCRFTETWLA 360
QY 468 GVVEPRFRDTPKXKFKLGDYDDPVMVLSYLERVEVQSPPLAAMATMARIGAEHVKASAM 527
DB 361 TKVPPDRFDLKERTRAPLQPFDDPDVQTYLDAIVEL----- 396
QY 528 QALQKVPKRTDMSAEPRKVOETVSVDP-----VGNNGRDSRPGVIAEAVRS 580
DB 397 -----PSDLMPTPLPVEBLEVRSSILAKELPTPATG-----VAP 432
QY 581 ENFETNDYAIRAGVSESSVDETVEMSVADMLKEASVYILAAGVAILISFQKRYFAS 640
DB 433 PRRRRRRSERPATAKRLPLPWIGLGV-----VVVLGGGCV-----WAMRS 475
QY 641 SSSSFORKDVMSSMSDVATIGSVADDSALP-----RMDARTANIVSKMKQIKS 691
DB 476 RSN-----STPPTPPPVQVTLPEAVPASPAPVAVALDRAQLETVLQWMLAKA 524
QY 692 LAEPEDRIEMLPVLDGRMLKIWTDRABETA--QLGLVYDYTLILKLSVDSVTYSADSTR 749
DB 525 AALGPQYDRDLAVLVTGEVLQTOGFSQQAANTQLSQFDH--KLTVDSVQLSDGQR 581
QY 750 ALVATLEESACLSDLVHPENNATDVRT--YTRYEYFMSKGMKLTGSGVL 799

DB 582 AVQAKVDE---VEQYVRGDDQLLETRRDGLGVLIRYQLVRENNIMKIASISLV 630

RESULT 14

ID ADJ38253 standard; protein; 819 AA.

AC ADJ38253;

DT 06-MAY-2004 (first entry)

DE Plastid division-related Arc6 orthologue protein 40.

XX prokaryotic type; plastid division; Ftn2; ARC6; ARCS; Fzo; plant cell;

KW agronomic; horticultural; crop plant; ornamental plant; woody plant;

XX herbicide target.

XX Arabidopsis thaliana.

PN WO2004001003-A2.

PD 31-DEC-2003.

PE 20-JUN-2003; 2003WO-US019536.

PR 20-JUN-2002; 2002US-0390140P.

PR 09-AUG-2002; 2002US-0402242P.

PR 20-JUN-2003; 2003US-00600070.

PA (UNMS) UNIV MICHIGAN STATE.

PI Oosteryoung KW, Vitha S, Kokeharova OA, Gao H;

DR WPI; 2004-082486/08.

DR N-PSDB; ADJ38252.

XX New isolated Ftn2, ARCS and/or Fzo-like nucleic acid sequences, useful

XX for further characterizing plastid division in plant cells, and in

XX varying agronomic and horticultural characteristics of economically

XX important plants.

XX Disclosure; Fig 8; 287pp; English.

XX This invention relates to novel prokaryotic type or plastid division and

XX related genes and proteins. In particular, the invention relates to novel

XX Ftn2 (ARC6), ARCS and Fzo-like genes and polypeptides. The methods and

XX compositions of the present invention are useful for further

XX characterizing plastid division in plant cells, in order to vary

XX agronomic and horticultural characteristics of economically important

XX plants, such as crop, ornamental and woody plants. They can also be used

XX as herbicide targets. The present sequence is that of a protein which is

XX related to the invention.

XX Sequence 819 AA;

Query Match 10.6%; Score 431; DB 8; Length 819;
Best Local Similarity 23.5%; Pred. No. 5e-29;
Matches 200; Conservative 116; Mismatches 311; Indels 224; Gaps 32;

QY 56 SSSSSPATATTTATVSLPESIDRPRHVPPIPIFYQVLAQTHTLTDGIRAFARVSK 115
DB 81 SRTSSLAASIST-----IELPVTCYQLIGVSEQAQKDEVKSVINLKT 124
QY 116 PPQGFSDDALISRQIIQAACETLSNPRSREYNEGLDDEATYITVPMDKVPALC 175
DB 125 DAEQGYTMEAAAQAQDILMDVRDL---LFESEYAGMLKEKIAKSPURIPWAWLPGALC 181
QY 176 VLOEGGETIYLRVGEALTKERLPKSKODVVMALAFDVSADAMALDPPFITGYEF 235
DB 182 LLOEVGGEKVLVDIGRAALRLDSKPYTHDIFLSMALBECAIARAAREVNVVS--QGFEA 239
QY 236 VEALAKULOEE-GASSIAPDIFRAQIDETLEITPRVYELGLPLGDDYAAKRLNGLSGV 294

```

Db      240 LARAQFLKSKVTLGKLA--LTQIEESLEGAPCTLDLGLPTPENARRGAIAL 297
QY      295 RNLMSVGGGAGALVGLTRK-----FMNEAFLEMTAEQVDF-----VATP 339
Db      298 RELLRQ-----GLSVEASCOIQDWPFCFLSQIISRLATEIVDLPMDDLATRK 346
QY      340 SNIPAESFE-----VEVALALVAQAFIGKKPHLLQDADKQFOLOQAKVMAEIP 390
Db      347 NKKSLSHNRGVVIDNCFYMWLGHIAVFGS-----KQNTINKAKTIGCECLI 396
QY      391 AMLYDRNMWEIDPGLERGLCALLIGKVDCEGRMWLG----DSEDQYRPAIVEPLEN 446
Db      397 A-----SEGVDLKKEBAFCFLKQGSBALEKLKQLESNSDSAVRNS-----ILGK 444
QY      447 SNRDDNDLGLCKLETWLAGVVPFRPRTK-----DKKFKGDDYDDPMVL 494
Db      445 ESKSTS-----ATPSLEAMLMESVLANFPDTRGCSPLANFPAEKKYIPNKKMGSSIM 499
QY      495 SY-----LERVEVQSSPLAAATWARIAGEHVKASAMQALOKVPSRYTDRNSAEPKD 548
Db      500 NHTNQRPILSTTQFVNS-----QHL-----YTAVEQLPTD 531
QY      549 VQETVPSVDVGNNGVGDGEPGVFIABAVRPSNFETNDYAIRAGVSESSV-DETTVENS 607
Db      532 LQSPVVSAAK--NNDE-----TSASMPVSQVKRN-----LGVHKNKIMDEW--LS 571
QY      608 VADMULKASVKTLAAGVAIGLISLFSQKY-----MESDVATIGSVRADSEALPR 673
Db      572 QSSLIGRVSVVAL-----LGCTVFFSLKSGINSGLQSPISVSARPHSESDSFLMKTE 626
QY      637 ---FLKSSSSFFORKDMSV-----MESDVATIGSVRADSEALPR 673
Db      627 SGNFRKNLDSVNNNGIVGNIKVLIDMLKMGHHPALYLKSSGQATSLSHSSELHKR 686
QY      674 -MDARTAEIVTSKQKIKSLAFEPDRHEMLPEYLDGRMLKIWTDRAEPLQGLV-DY 731
Db      687 PMDTEEBEELVRQWENYKABALGPTHQVYSLSEVLDESMLVOW-QTLAQTAEAKSCYMRP 745
QY      732 TLTKLSDSVTVSAD--GTRALVEATLESACSLDVHPENNATDVRVTTTREVFWMSK 788
Db      746 VLAHLEVLQAHTEBDGAGBAETELAEAEELVDESQPK-NKXYSTYKIRIYILKQOE 804
QY      789 SG-WKITEGSV 798
Db      805 DGLMKFCQSDI 815

RESULT 15
ADJ38251
ID      ADJ38251 standard; protein; 819 AA.
XX      AC      ADJ38251;
XX      DT      06-MAY-2004 (first entry)
XX      DE      Placid division-related Arc6 orthologue protein 39.
XX      KW      prokaryotic type; placid division; Ftn2; ARC6; ARC5; Fzo; plant cell;
XX      herbicide target.
XX      OS      Arabidopsis thaliana.
XX      PN      MO2004001003-A2.
XX      PD      31-DEC-2003.
XX      PF      20-JUN-2003; 2003MO-US019536.
XX      PR      20-JUN-2002; 2002JUS-0390140P.
XX      PR      09-AUG-2002; 2002JUS-0402242P.
XX      PR      20-JUN-2003; 2003JUS-00600070.
XX

```

```

PA      (UNMS ) UNIV MICHIGAN STATE.
XX
PI      Oosteryoung KM, Viltha S, Kosharova OA, Gao H;
XX
DR      WPI; 2004-082486/08.
XX
PT      New isolated Ftn2, ARC5 and/or Fzo-like nucleic acid sequences, useful
PT      for further characterizing placid division in plant cells, and in
PT      varying agronomic and horticultural characteristics of economically
PT      important plants.
XX
PS      Disclosure; Fig 8; 287bp; English.
XX
CC      This invention relates to novel prokaryotic type or placid division and
CC      related genes and proteins. In particular, the invention relates to novel
CC      Ftn2 (ARC6), ARC5 and Fzo-like genes and polypeptides. The methods and
CC      compositions of the present invention are useful for further
CC      characterizing placid division in plant cells, in order to vary
CC      agronomic and horticultural characteristics of economically important
CC      plants, such as crop, ornamental and woody plants. They can also be used
CC      as herbicide targets. The present sequence is that of a protein which is
CC      related to the invention.
XX
SQ      Sequence 819 AA;
XX
Query Match          10.6%; Score 431; DB 8; Length 819;
Best Local Similarity 23.5%; Pred. No. 5e-29;
Matches 200; Conservative 116; Mismatches 311; Indels 224; Gaps 32;

QY      56 SSSSFATATTATVLSLPPSIDRPHVPIPIIDFQVLAGQTHFLTDGIRRAFEARVSK 115
Db      81 SRTSSLAASTST-----LELPVTCYOLGVSRQAEKDEEVKSVINLTKT 124
QY      116 PPGGFSDDALISRRQILOACETLSNPSRRRYNGSLDDEEATVITDVPWKGALC 175
Db      125 DAEGYTMELAAARQDLMDVDRKL--LFESYAGNLKXIAPKBPLRPMMLEALC 181
QY      176 VLOGGETEIVLAVGALTKERLPKSFQDQVVLVMAIAFDVSRDAMALDPDFTGYEF 235
Db      182 LLOVGQEKLVLDIGRAALRNDSKPYIHDIFLSMALBGAIKAAFEVVKV--QGFBA 239
QY      236 VEBALKLOEE-GASSLAPDLRAQIDETLETPRYVELLGLPLGDVYAAKNGLSGV 294
Db      240 LARAQFLKSKVTLGKLA--LTQIEESLEGAPCTLDLGLPTPENARRGAIAL 297
QY      295 RNLMSVGGGAGALVGLTRK-----FMNEAFLEMTAEQVDF-----VATP 339
Db      298 RELLRQ-----GLSVEASCOIQDWPFCFLSQIISRLATEIVDLPMDDLATRK 346
QY      340 SNIPAESFE-----VEVALALVAQAFIGKKPHLLQDADKQFOLOQAKVMAEIP 390
Db      347 NKKSLSHNRGVVIDNCFYMWLGHIAVFGS-----KQNTINKAKTIGCECLI 396
QY      391 AMLYDRNMWEIDPGLERGLCALLIGKVDCEGRMWLG----DSEDQYRPAIVEPLEN 446
Db      397 A-----SEGVDLKKEBAFCFLKQGSBALEKLKQLESNSDSAVRNS-----ILGK 444
QY      447 SNRDDNDLGLCKLETWLAGVVPFRPRTK-----DKKFKGDDYDDPMVL 494
Db      445 ESKSTS-----ATPSLEAMLMESVLANFPDTRGCSPLANFPAEKKYIPNKKMGSSIM 499
QY      495 SY-----LERVEVQSSPLAAATWARIAGEHVKASAMQALOKVPSRYTDRNSAEPKD 548
Db      500 NHTNQRPILSTTQFVNS-----QHL-----YTAVEQLPTD 531
QY      549 VQETVPSVDVGNNGVGDGEPGVFIABAVRPSNFETNDYAIRAGVSESSV-DETTVENS 607
Db      532 LQSPVVSAAK--NNDE-----TSASMPVSQVKRN-----LGVHKNKIMDEW--LS 571
QY      608 VADMULKASVKTLAAGVAIGLISLFSQKY-----MESDVATIGSVRADSEALPR 673
Db      572 QSSLIGRVSVVAL-----LGCTVFFSLKSGINSGLQSPISVSARPHSESDSFLMKTE 626

```

```

QY 637 ---FLKSSSFQRKDWSS-----MESDVATIGSVRADDSBALPR 673
Db 627 SGNFRKRLDSDVNRNGIYGNIKVLDMLKMHGHPDALYKSSQSATSLSHSASBLHKR 686
QY 674 -MDARTANIVSKWQKIKSLAFSGPDHRIEMLPEVLDGRMLKIWTDRAAETAOGLVY-DY 731
Db 687 PMOTEEABEIVRQWENYKABALGPTHQVYSLSEVLDESMLVQW-QTLAQTAEAKSCYWRP 745
QY 732 TLKLSYDSVTVSAD---GTRALYEATIESACLSDLVHPENNATDVRTYTTTRYEVFWSK 788
Db 746 VLLHLEVLQAHIFEDGIAGEAAIEALIEAAELVDESQPK-NAKYYSTYKIRYILKKOE 804
QY 789 SG--WKITEGSV 798
Db 805 DGLMKFCOSDI 815

```

Search completed: June 10, 2005, 01:38:50
 Job time : 101 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2005, 01:02:57 ; Search time 28 Seconds
(without alignments)
2135.496 Million cell updates/sec

Title: US-10-600-070-2

Perfect score: 4063
Sequence: 1 MEALSHVIGIGLSPQLCRLP.....YEVWMSKSGWKITRGVTLAS 801

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145.5	3.6	6095	US-09-144-085-2	Sequence 2, Appl
2	143	3.5	2089	US-08-418-893D-23	Sequence 23, Appl
3	143	3.5	2089	US-08-418-893D-24	Sequence 24, Appl
4	139.5	3.4	954	US-09-057-969-2	Sequence 2, Appl
5	131.5	3.2	1194	US-08-680-326-35	Sequence 35, Appl
6	128.5	3.2	876	US-08-436-664-20	Sequence 32, Appl
7	128.5	3.2	876	US-08-436-664-32	Sequence 32, Appl
8	128.5	3.2	876	US-08-436-664-34	Sequence 34, Appl
9	128.5	3.2	876	US-09-135-642-20	Sequence 20, Appl
10	128.5	3.2	876	US-09-135-642-32	Sequence 32, Appl
11	128.5	3.2	876	US-09-135-642-34	Sequence 34, Appl
12	128.5	3.2	876	US-08-394-232A-20	Sequence 20, Appl
13	128.5	3.2	876	US-08-394-232A-32	Sequence 32, Appl
14	128.5	3.2	876	US-08-394-232A-34	Sequence 34, Appl
15	128.5	3.2	876	PCT-US95-04080-20	Sequence 20, Appl
16	128.5	3.2	876	PCT-US95-04080-32	Sequence 32, Appl
17	128.5	3.2	876	PCT-US95-04080-34	Sequence 34, Appl
18	126.5	3.1	2482	US-09-252-991A-16967	Sequence 16967, A
19	125.5	3.1	5087	US-09-144-085-1	Sequence 1, Appl
20	122.5	3.0	2756	US-08-375-709-11	Sequence 11, Appl
21	122.5	3.0	2756	US-08-753-929-11	Sequence 11, Appl
22	122.5	3.0	2756	US-09-090-793-7	Sequence 7, Appl
23	122.5	3.0	2756	US-09-231-899-7	Sequence 7, Appl
24	121.5	3.0	680	US-09-252-991A-26639	Sequence 26639, A
25	121.5	3.0	947	US-09-540-236-1991	Sequence 1991, Ap
26	121.5	3.0	1220	US-09-540-236-3011	Sequence 3011, Ap
27	120	3.0	852	US-09-585-858-19	Sequence 19, Appl

28	120	3.0	852	4	US-10-270-878-19	Sequence 19, Appl
29	120	3.0	1262	4	US-09-198-452A-97	Sequence 97, Appl
30	120	3.0	1266	4	US-09-438-185A-83	Sequence 83, Appl
31	119.5	2.9	831	1	US-08-073-384C-5	Sequence 5, Appl
32	119.5	2.9	831	1	US-08-254-359A-5	Sequence 5, Appl
33	119.5	2.9	831	1	US-08-463-043-5	Sequence 5, Appl
34	119.5	2.9	831	1	US-08-481-238-5	Sequence 5, Appl
35	119.5	2.9	831	2	US-08-471-066B-5	Sequence 5, Appl
36	119.5	2.9	831	2	US-08-484-956-5	Sequence 5, Appl
37	119.5	2.9	831	2	US-08-757-653-5	Sequence 5, Appl
38	119.5	2.9	831	2	US-08-559-491-5	Sequence 5, Appl
39	119.5	2.9	831	2	US-08-756-386-5	Sequence 5, Appl
40	119.5	2.9	831	2	US-08-823-516-5	Sequence 5, Appl
41	119.5	2.9	831	3	US-08-682-853A-5	Sequence 5, Appl
42	119.5	2.9	831	3	US-08-759-038-5	Sequence 5, Appl
43	119.5	2.9	831	3	US-08-758-314-5	Sequence 5, Appl
44	119.5	2.9	831	3	US-09-350-309-5	Sequence 5, Appl
45	119.5	2.9	831	3	US-08-520-946-5	Sequence 5, Appl

ALIGNMENTS

```
RESULT 1
US-09-144-085-2
Sequence 2, Application US/09144085
Patent No. 6280999
GENERAL INFORMATION:
APPLICANT: Gustafsson, Claes
APPLICANT: Betlach, Mary C.
APPLICANT: Ashley, Gary
APPLICANT: Tulien, Bryan
APPLICANT: Ziemann, Rainer
TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
FILE REFERENCE: 30062-20020.20
CURRENT APPLICATION NUMBER: US/09/144, 085
CURRENT FILING DATE: 1998-08-31
EARLIER APPLICATION NUMBER: 09/010, 809
EARLIER FILING DATE: 1998-01-22
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 6095
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-144-085-2

Query Match          3.6% Score 145.5; DB 3; Length 6095;
Best Local Similarity 19.3%; Pred. No. 0.0034;
Matches 167; Conservative 128; Mismatches 349; Indels 221; Gaps 38;

70 LVSLPSPISDRPERH-VPIPIFYGVLAGQHTFLTDGIRAEARVSKPPQGFSDALIS 128
1642 LALIDALSRREALIVVHLDLAQ-----LQGLSGSGELPALF-----RALTL 1684
129 RROTLQACETLSNPRSRREYNEGLDDEKATVTVDPMDVPALCVLQGGETEYVLR 188
1685 RPSLRKKSASATRRPASALRERLSALPEAERLNALVEVRGEV-AAVAGLQK----- 1734
189 VGEALTLERLPKSKQDVVLVLMALAFDVSRDAMALPPDPITGYEFVEBALKLLQEGA 248
1735 -GEAVAADQVKEIGLDSILMAVLRNLRTSKRTSL-PATIVFVYPPPRATIELLKQAF 1792
249 SSL-APLRAQIDETLEITPRVYVLEL-----GLPLGDDYAAKRLNGLSGVNRI--LWS 300
1793 SGLQVKEARARVRRRAKDEPIALVSMACRLPGVAVAPPDYWRLLAGKQALIEGLPARWD 1852
301 -----VGGGASALVGGLTREKPMNER-----FLRMTAAEGVD 333
1853 GFEYVDDPEPAAGKSVAREGGFVADIDLFDANFGISPREAOSMDPOHRLVLETAWALE 1912
334 LVFATPENIPAESFEVYEVALLA-----LVAQAFIGKKPHLLQDADKQFOQLQ 380
```

```

Db      1913 RAGVRPALSGSAGVYIGSGSDYGAHITVDLKEELGYRIGISAASISG-----R 1964
Qy      381 QAKVMAEITPAMLYDTNNMEIDFGLERGLCALLIGKVEDECRMWIGLDSBDSQYRNPAL- 439
Db      1965 VAYALGIQGRPMYTVDICSSSL-VSLHACTALNQGBD-----LALAGYVWSTPLLF 2018
Qy      440 VEFV-LENSNRDNDLPGCLKLET-----WLAG--VVPFRPTDKCKFKLGDYD 489
Db      2019 VEFSLKGMGRD-----GRCKSFSVQADAGMAGCGMILLKTLSDAQ-----RD 2063
Qy      490 DPMVLSL-----EKEVYVQ-----504
Db      2064 GDRVLGVRGSAAVQDSQGLTPANPAGQVRVIALSSGSLSPEDIDAVEAHGTGTSL 2123
Qy      505 GSPPLAAATATARGAE-----HYKASAMQA--LOKVPFSRYDRNSAAE 545
Db      2124 GDPIEAGALAEVQRPSPERPLYLSSSKSNLGHAAQAAVAGVYIKVYL-----SMQ 2175
Qy      546 PKDVQETVFSVDPVGNVNGRDEBPVFIAEAVRSENFETNDYAIRAGVSSSVDETTVE 605
Db      2176 HEVLPKTLHAEP-SPHIGWEGS-GLSLQEARP---WRNRGRVRAVGSFGISGTNAH 2230
Qy      606 MSVADMLKEA---SVKLLAGVAIGLISLPSQKTFKSSSSFOKDMVSSME---SDYAT 659
Db      2231 IILEEAPAEARREBEVEEAAPALLPLVLSGRDEAAVNAQAQRMKWLBEHGEVGSMDVVR 2290
Qy      660 IGSVRADDSALPRMDARTAEINIVSKWQIKSLAFG-PDHRIBMLPEVLDGRMLKIWTR 718
Db      2291 TAAHHRHFSRSRAVLAASAGAV---EGLRALSSGRPDAAVSGTAKRGKLAVALTTGQ 2347
Qy      719 AETAQAQ--LVYDYTLKLKLSVDSVTVSAD-----GTRALVEATL--EESACT--SDLVH 767
Db      2348 GSGRLGKSKLVEYVYVFAAFDEVCEALDAYLDRGAREVVFALAAAGEGALLERTETGQ 2407
Qy      768 PENNAIDVYRTTYTRIVYFWSKSGWK 792
Db      2408 PGLFALEVALYRQ-----WESWGLK 2427

RESULT 2
US-08-418-893D-23
Sequence 23, Application US/08418893D
Patent No. 5559220
GENERAL INFORMATION:
APPLICANT: ROESSLER, PAUL G
APPLICANT: OHLROGGE, JOHN B
TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTETILLA CRYPTICA
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
STREET: 1617 Cole Blvd.
CITY: Golden
STATE: CO
COUNTRY: USA
ZIP: 80401-3393
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/418,893D
FILING DATE: April 7, 1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/104,938
FILING DATE: September 14, 1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: O'CONNOR, EDNA
REGISTRATION NUMBER: 29,252

```

```

; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
;
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-418-893D-23

Query Match 3.5%; Score 143; DB 1; Length 2089;
Best Local Similarity 20.0%; Pred. No. 0.0009; Indels 272; Gaps 44;
Matches 169; Conservative 113; Mismatches 289;

Qy      73 LPSIDRPERHVPIDIFYVLAQTHFTLDGIRRAFEARVSKPP-----QFGFSDAL 126
Db      197 LPNALDK-----LGIFIGPTGPVMSVLGDKLAANILAQTAVPSPISWGSFGGPDGP 250
Qy      127 ISRRQIIQAACETLSNPRSREYNEGILD--DEBAYITTPVMDKVGALCVLOEGG--- 181
Db      251 IQ-----ADLTBEETGTPMEIFNKGLVTSADEAVIYANKIGMEN--GIMIKASEGGGCK 301
Qy      182 -----ETEIYLR-----YGEALLKERLPSFKQDVVLWALAFVLSRDNALDP 226
Db      302 GIRFVDEADLRNAFVQVSNVEIGSPIFLMQLCKNAHIVQVQV---DQGNAYALNG 357
Qy      227 PDFITGEFEVEBALKLOEBGASLAPDLRAQIDETLEETTPRVVLELGLPLGDDVYAK 286
Db      358 RDCSTGRF-----QKIFEEGPPSIYP-----KETFHEM-----ELAAQ 391
Qy      287 RLNLGSGVRNIIILSVGGGASALVGLTRKFPNMEAFRLRTAAEQVDLFVA---TPSNIP 343
Db      392 RL-----TONIGYO--GAGTVEYLYNNAADNKF--FLEINPLQYEHVPTEGITGNLP 441
Qy      344 AES-----FEVVEVALALVAQAFIGKKPHLLQADQFOFOLOQAQVMAEIPA--- 391
Db      442 ATQLQVAMGIPLFNIPIRLYGREDAYGTP--IDFLQERYEL--DSHVIYAIRITENP 498
Qy      392 -----MYDTRNNW-EIDFGLERGLCALLIGKVEDECRMWIGLDSBDSQY-- 434
Db      499 DEBPKPTSGSIEIKPGSTPNWGVFSVGANGI-----HEFADSGFGH 542
Qy      435 ---RNP-----AIVEFVLENSNRDD-NDLPGCLKLET-----WLAGVVP 472
Db      543 LPAKGPMBEQARRKALVIALKEMEVRGDIRNSVEYLVLLETEAFKNTTIDTWSLDGII-- 600
Qy      473 RFRDIDDKRFKLDYDDPMVLSYLERVEVYQSPPLAAATMARIGAEHFKASAMQA--- 529
Db      601 ---KEKSVKV-----EMPSHL-----VVGAAGAFKAEHVKAVATEEVESFRKQVS 644
Qy      530 -----LOKVPSRYTDRNSAEPKDVQET-----552
Db      645 TAGIPGINSFNIENVAYLDTKYPHVERISPDVYRFLTLDGTTIVEVYQTAEGALLATFGG 704
Qy      553 ---VPSVD-PVGNNGVRGE---PGVFIAEVRBS---ENFETNDYAIRAGVSES 597
Db      705 ETRHIFGMDPEPLRLSLDQATVLMPTIFPSELRKTDVTGKVVRVYLDQNGATYEAQ--- 760
Qy      598 SVDETTVMSVADMKEASVKIILAAGVAILGLISLFSQKYFLKSSSFQKDMVSSME-SD 656
Db      761 ---QPYVEVAMKMT--MPKATESG-----KITHNLASGVYSAGDLASLELKD 806
Qy      657 VATIGSV-----RADSEALPRMDART--ENIVSKWQIKSLAFGPDHRIEMPEVLDGR 710
Db      807 PSRVKKIETFSGLDDMESKVDLEPKAVANNVLS-----GFNLD-----PEAAVQ 852
Qy      711 MKIWTDRAAETAQLGLVYD-VYTLKLKLSVDSVTVSADGTRALVEATLEESACLSDLVHP 769

```

Db 853 AIDSATSSAAADLLVQVLDDEFYVESQFDGV-IADVVRILTANTE-----TLDVISE 907
Qy 770 NNA 772
Db 908 NLA 910

RESULT 3
US-08-418-893D-24
/ Sequence 24, Application US/08418893D
/ Patent No. 5559220
/ GENERAL INFORMATION:
/ APPLICANT: ROESSLER, PAUL G
/ APPLICANT: OHLENGER, JOHN B
/ TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
/ TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
/ NUMBER OF SEQUENCES: 25
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: NATIONAL, RENEMABLE ENERGY LABORATORY
/ STREET: 1617 Cole Blvd.
/ CITY: Golden
/ STATE: CO
/ COUNTRY: USA
/ ZIP: 80401-3393
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/418,893D
/ FILING DATE: April 7, 1995
/ CLASSIFICATION: 800
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/104,938
/ FILING DATE: September 14, 1993
/ CLASSIFICATION: 800
/ TELEPHONE: 303-231-1000
/ TELEFAX: 303-231-1098
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 24:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2089 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULAR TYPE: protein
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: N-terminal
/ US-08-418-893D-24

Query Match 3.5%; Score 143; DB 1; Length 2089;
Best local Similarity 20.0%; Pred. No. 0.0009;
Matches 169; Conservative 113; Mismatches 289; Indels 272; Gaps 44;

Qy 73 LPPSIDRRHVPIDFYQVLAQTHFLNDGIRAFARVSKP-----QGFPSDAL 126
Db 197 LFNALDK-----GKIFGPTGPMVSVDKIANIILAAQAKVPSIPWSSGFGPDGPF 250
Qy 127 ISRQOIIQAACETLSNRSRREVEGLD--DEBAYVTDVPMKVPALCVLQEG--- 181
Db 251 LQ-----ADLTEGTIPMEIFPNKGLVTSADNAVYANKIGWEN--GIMIKASBGGGCK 301
Qy 182 -----ETEVLR-----VGEALIKERLPKSFQDVVLAVALFLVDSRDAMALDP 226
Db 302 GIRFVDKADLRNMFVQVSNVEVIGSPIFLMQLCKNARIIEVQIVG---DQHGNAVALNG 357

Qy 227 PDFITGEFEVEALKLLOEBGASLADPLRAQIDETLEETPRVYLLGLPLGDVYAAK 286
Db 358 RDCSTORF-----QKIFEGSPSIVP-----KETHHEM-----ELAAQ 391
Qy 287 RLNLGSGVRNLIWVGSGASALVGLTREKFNNEAFLRMTAQVQDLFVA--TPSNIP 343
Db 392 RL-----TONIGVQ--GAGTVEYLYNADNKF--FLEINPRLOVHEHPVTEGITGANLP 441
Qy 344 AES-----FEVEVALALVAQAFIGKKPHLLQDADQFQOLQOAKVAMAEIPA--- 391
Db 442 ATOLQAMGIPLEFNIPIRLYGRDAGYDPE--IDFLQERREL--DSHVIABITLENP 498
Qy 392 -----MLYTRNM--EIDFGLERGCALLIGKVCRCMWLGDSBSQY-- 434
Db 499 DEGRKPTSGSERIKFPOSTRWVWGYFSVGANGI-----HEFADSQGH 542
Qy 435 ---RNP-----AIVEFLSNSRDD--NDLPGLCKLET-----WLAGVFP 472
Db 543 LFAKGPMBQARKALVIALKEMEVRGDIIRNSVEYLVKLTETEARFKNTIDTSMLDGIT-- 600
Qy 473 RFRDTKCKKRLGDYDDPMVLTSLERVEYVQSPPLAAATMAIGAEHVYASAMQ--- 529
Db 601 ---KEKSVKV-----EMPSHL-----VVVGAIVKAFERHVEVATEEVEKSEFRKQVS 644
Qy 530 -----LOKVPERSYTRDSABPKVOET----- 552
Db 645 TAGIPGINSRNIEVAYIDTKPFVVERISPDVYRFTLDGNTIDVEVQYTAAGALLATRG 704
Qy 553 ---VPSVD--PVGNNGVRDGE---PGVFLAEAVRPS-----BNFTNDYAIRAGVSES 597
Db 705 ETRHIFGMDDELGLRLDLGATVLMPTIFDPSBLRTVTKGVVYLDQNGATVAG--- 760
Qy 598 SVDETYEMSVADMLKASVKIILAGVAILGISPSQKTFKSSSSFORQDMVSMQ--SD 656
Db 761 ---QPYVEVEAMKMI--MPIKATESG-----KITHNLSAGSVISAGDLLSLBKD 806
Qy 657 VATIGSV-----RADSEALPRMDART--ENIVSKWKIKSLAGPHEIMLEPVLDGR 710
Db 807 PSRVKIKETTSCKDIDNESKVDLEPQAVMNVLS-----GNLD-----PEAVAOQ 852
Qy 711 MLKIWTDRAAETAGLVYD--YTLIKLSDSVTVSADGTRALVEATLESACSLDVHPE 769
Db 853 AIDSATSSAAADLLVQVLDDEFYVESQFDGV-IADVVRILTANTE-----TLDVISE 907
Qy 770 NNA 772
Db 908 NLA 910

RESULT 4
US-09-057-969-2
/ Sequence 2, Application US/09057969
/ Patent No. 6013451
/ GENERAL INFORMATION:
/ APPLICANT: WONG, VICTOR THI WONG
/ APPLICANT: PHANG, SENG MENG
/ TITLE OF INVENTION: BACILLUS STEAROTHERMOPHILUS DNA
/ TITLE OF INVENTION: POLYMERASE I (KLENOW) CLONES INCLUDING THOSE REDUCED TO
/ TITLE OF INVENTION: 3'-TO-5' EXONUCLEASE ACTIVITY
/ NUMBER OF SEQUENCES: 17
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
/ STREET: 745 FIFTH AVENUE
/ CITY: NEW YORK
/ STATE: NEW YORK
/ COUNTRY: USA
/ ZIP: 10151
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS

```

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/057,969
FILING DATE: 09-APR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KOWALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 674507-2001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: amino acid
US-09-057-969-2

```

Query Match 3.4%; Score 139.5; DB 3; Length 954;

Best Local Similarity 22.0%; Pred. No. 0.0005;

Matches 175; Conservative 132; Mismatches 315; Indels 173; Gaps 44;

```

99 HPLTDGIRRAFEARVSKPQFGFSDALISRQIIOACETLSNPSRREY-NEGLLDDE 157
183 HYEADDIIGMAARAERE--GFAVKYISGDRDLTQ-----LASPQYVEITKKGIDIE 234
158 EAVYITDY-----PMDKVPG-----ALCVQEGEGETIVLRVG 190
235 SYPEYVEYKGYLTPEQIVDLKGLMGDKSDNIPGVPGIGKTKVAKLKQFQVENVLASI 294
191 EALKERLPKSPKO--DVLVMAALFDVSRDA--MALPDPDITGVGEFEALKTLOE 245
295 DEIKGEKTKENLRKYRDLAL-LSKOLAICDPAVEYLTLD--DIVYGEDEKVKVALLFOE 351
246 EGASIALPDLRAQIDE-----TLEETPRVYLELGLT--PLGGDYAAKRLNG- 290
352 LGFQSFIDKMAVQDEBEKPLAGMDPAIDSVTDEMLADKAAALVEVVGVDYHNAPIVGI 411
291 -LGGVRIIL-----SYGGGASALVGGTLREKFM-----NEAFLEMTAEQV-D 333
412 ALNERGRFLRPETAADVPKFLMLGDETKKTMFDSKRAVALNKGIEIAGVGVPFD 471
334 LFVATPSNIPAES-----FEVVEVALVAQAFIGK-KPHLLQDAKQPOOL--QOAK 383
472 LLLAAVILLDPQAQGDVAAVAKMAYEAVNSDEAVYGGAKRTVPDEPTLAEOLVKRAAA 531
384 VMAMEIPAMLYDTRNNWE-IDFGLEKGLCALLIG-----KVDECRM-WLGLD--SEDSQY 434
532 IWLAEPLMELRNEDDRLLTELEHNLAGILANMEFTGVVVDKRLKQMGAEITQLO- 590
435 RNPALYEFVLNSNRDNDLPGICLLLETWLAGVPPRRFDITOKKPKGLGDIYDDPMVL 494
591 ---AVERRIYELGQEFNINSP--KQIGT---VLDPKQLPVLKTKTKG--VSTSDVL 639
495 SYL-BREVVVGSLAAMATMARIGAEHVAKSAMQALQKVPKSYTRNSAPRDVQET- 552
640 EKLAIPHHEIYE--HILHYRQIGKQSTYIE-GLLKVVHPVTKVHTFMQA---LQOTG 692
553 -VESVDPVGNVNGRDESPGVFIAEAVRPS-----NPEFDYAIRAGVSESSVDETT 603
693 RLSSVEENLQNIPIRLBEGKRIRQAFVPSBDMLIFADYSQIELRVLAHIAE---DONL 749
604 VE-----NSVADMLKE--ASVKIILAGVAILGISLFSQKFKYFKSSSSSQRK 647
750 IEAFRRWLDIHTKTAMDIHFVSEEDVTANMRQAKAVFGIIVGISD-YGLAQMLNITRK 808
648 DMSVSMESVAVTIGSVADSEBALPRMDARPAENIVKWKOKISLAGPDHRIEMLEVL 707
809 EAABEFIRYRASPFGVK-----QYMDNIVQE-AKQKGYVTTLHRRRYLDPIT 855

```

```

QY 708 DGM-LKIWTDRRA-ETAQLGLVYDYTLKLSVD-SYTVSADGTRA-----LVEA 754
DB 856 SRNFVNTFAERFAMTMTPIQGSAD-IKKAMWIDLSVSVREERLQALLLQGHDELLEA 914
QY 755 TLESACLSLVHPE 769
DB 915 PKEBIGRLCRLV-PE 928

```

RESULT 5

```

US-08-680-326-35
Sequence 35, Application US/08680326
Patent No. 5925733
GENERAL INFORMATION:
APPLICANT: ROSE, TIMOTHY M.
APPLICANT: BOSCH, MARNIX
APPLICANT: STRAND, KURT
APPLICANT: TODARO, GEORGE J.
TITLE OF INVENTION: DNA POLYMERASE OF GAMMA HERPES VIRUSES
TITLE OF INVENTION: ASSOCIATED WITH KAPOSI'S SARCOMA AND RETROPERITONEAL
NUMBER OF SEQUENCES: 152
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,326
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Schiff, J. Michael
REGISTRATION NUMBER: 40,253
REFERENCE/DOCKET NUMBER: 29938-20001.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 1194 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-680-326-35

```

Query Match 3.2%; Score 131.5; DB 2; Length 1194;

Best Local Similarity 19.0%; Pred. No. 0.0046;

Matches 168; Conservative 122; Mismatches 297; Indels 297; Gaps 44;

```

QY 48 SDNFNFSRSS-----SSPATTTTATIVSLPSIDRPERHY--PIPIDFYVLGAQ--TH 99
DB 205 NDATLNGDKNAFHTGSKSASPSFRVEV--TERTDVYVYDTQPCAFYVYBSSSFTN 261
100 FLTDGIR---RAFEARVSKPQFGFSDALIS-----RRQI----- 132
262 YLCDNPHPELAKYEGRDATTRFLMNDPGFVSGWYQKRGVDGERVRVAPASRQLTSD 321
133 LQAAC-----ETLSNPSRREYNEGGLIDE-----EATYITDVPMDKVPGALC 175
322 VEIDCMSDNLQAIPLNDSDWPDYKLLCFDIECKSGSGSNELAPFPAITHLEDL----VIQISC 377
176 VL-----QEGEGETIVLRVGKALKEKRLPKSPKO-----DVLVMAALFDVSRD 220
378 LTVSIPQSLHILLPSLSCDLPORYVQEMKAGLPEPTVLEBDFELLINFMILVKQ 437

```

```

Qy 221 AMALDPDPTITGYEVEBEALKLOEGASLAPDLRAQIDETLEETPRVYLLGLPLG 280
Db 438 Y-----AEFATGVNIIVPDAFI-----MEKINSYSLK----- 468
Qy 281 DDYAAKRLNGISGVRLTMSVGGGASALVGLTREKFMNEAFIRMRAEVDLFPATPS 340
Db 469 DGYSISIRNGGLFKI-----MDVKGSG-----FORSKVKINGLISLDMVALATE 512
Qy 341 NIPAESEVEVEA-----LATVAQAFIG 363
Db 513 KKLKSSYKLSVAAEALNESKRDLPYKDIPEGYVAGPENTRGIIIGBYCTQDSALYGLKLPFK 572
Qy 364 KKPHLLQDADKQFOOLQOAKYMAEIPAMLYDTRNNMEIDGLERGLCALLIGKVBECRM 423
Db 573 YLPHLELSA-----VARLARITITRAIYD-----GQVRIYTCGLGLASS--- 612
Qy 424 WLGLDSEDSOYRNDPAIVEP-----VLENSNRDNDLPGCLKLETWLAGV-VFPR 473
Db 613 -RGFILPDGGY--PATEYKQVIPDVGPEBEDESVSPGTSSGRNVGKARVPD- 668
Qy 474 FPDYKDKKFKLGDYDDPMVLSYERV-EVYQSSPLAAATMARI GAHVYKASMOALQK 532
Db 669 -PDT-----GFYIDPVVYLDFASLYPSIIQAHL--CFTTLTLPETVK-----R 710
Qy 533 VPFERYTDRNSAEPRDQVETVFSYDP-----VGNVNGRDEGPVFIAE--AVRSENEET 585
Db 711 LNPEDY-----ATFTVGGKRLFFVRNSV-RESLGLVLLDKMLAMRK----- 750
Qy 586 NDVAIRAGVSESVDETTVMSVADMLKEASVKILA-----AGVAIGLIS----- 630
Db 751 -AIRARIPSSSDE-----AVLLDKQOAIKVVCNSVGFYGAQFLCLVVAATVT 802
Qy 631 -----LFGSKYTKLKSSSFQR-----KDMVSSMESVA-TIGSVRADSEALPRNDAR 677
Db 803 TIGRQMLSTEDYIHNWMAAFERITAPDIESSVLSQKAYEVKVIYGDTSVFIRFGKV 862
Qy 678 TAENIVSKMOK-----IKSLAFGPDHRIEMLPVLDGRMLKITWRAATAQGLVYDYL 733
Db 863 SVEIATIGERMAHITSTALFCFPKIECEKTFI-KLLI-----TKKKTIGVIYGGKV 915
Qy 734 LKLSVDSVTVS-----ADGTRALVEATL-----EESACLS 763
Db 916 LMGVVDLVRKNQCQFINDYARKLVELLLYDPTVSRAAAEASCVS 959

```

RESULT 6
 US-08-436-664-20
 ; Sequence 20, Application US/0843664
 ; Patient No. 5874282
 ; GENERAL INFORMATION:
 ; APPLICANT: RIGGS, MICHAEL G.
 ; APPLICANT: SIVARAM, MATTHOR
 ; APPLICANT: TUDOR, STALIA D.
 ; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
 ; TITLE OF INVENTION: STEROTHERMOPHILUS
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Gen-Probe Incorporated
 ; STREET: 9880 Campus Point Drive
 ; CITY: San Diego
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92121
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: PaateSeq Version 1.5
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/436, 664
 ; FILING DATE: 08-MAY-1995
 ; CLASSIFICATION: 536

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/394,232
FILING DATE:
APPLICATION NUMBER: 08/307,410
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: 08/222,612
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fisher, Carlos A
REGISTRATION NUMBER: 36,510
REFERENCE/DOCKET NUMBER: GP94003.CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-7929
TELEFAX: 619-546-7929
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-436-664-20

Query Match 3.2%; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0053;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

Qy 99 HFLTDGIRBAEAVSKPQGFSDDALISRRQILQACETLSNPRSREX-NEGILDE 157
Db 107 HYENDDIIGTMAABAEK--GFAVKYISGRDLTQ-----LAPQYVETLKKGIDIE 158
Qy 158 EATVITDV-----PMDKVPG-----ALCVLQEGGETEIVLRVG 190
Db 159 SYTETVEVVEKYGTLPEQIVDLKGMGDKSDNIPGVPIGKETAVKLKGFTVENVLASI 218
Qy 191 EALKERLPSFKQ--DVLVMAALFDVSRDA--MALDPDPTITGEFEBALKLOE 245
Db 219 DEIRGEKIKENTLRQYRDLAL-LSKQLAAICRDAPVELTLD--DIVYKGEDEKVALAFOE 275
Qy 246 EGASSLAPDLRAQID-----TLEETPRVYLLGL-----PLGDPYAKRLNG- 290
Db 276 LGFQSFIDKNAVQTBEGEKLPGAMDFAIDSVTDEMLADKALVVEVGVGDYHNHAPIVGI 335
Qy 291 -----LSGVNIIILSVGGGASALVGLTREKFMNEAFIRMTAAE----- 330
Db 336 ALANERGRFPLRPETALADPKFLAM-----LDDETKKTMFPGSKRAAVALLKWKGI 385
Qy 331 -----QVDLFPATPSNIPAS-----FEYEVALLVAQAFIGK-----KPHLLQ 370
Db 386 ELRGVDFDLILAAIYLDPAQAAGDVAAVAKNHQYEAVSDEAVVGKAKRTVPDEPTLAE 445
Qy 371 DADQFOOLQOAKYMAEIPAMLYDTRNNME-IDFGLERGLCALLIG-----KYDECRM 423
Db 446 HLAH-----KAAALWALBEPMLDELRRNEQDRLLTLEBQPLAGILANMEFGVKVDYRL 500
Qy 424 -WLGLD--SEDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVVPFRPDYDKK 481
Db 501 EQMGAELETLQIQ-----AVERRIYELAOGEFVINSR--KQIGT-----VLFKQLDPLVKK 549
Qy 482 PKLGDYDDPMVLSY-ERVEVQSSPLAAATMARI GAHVYKASMOALQKVPFRYTD 540
Db 550 TKTG-YSTADVLKELAPHEIIVE--HILHYRQLGKLQSTYIB-GLKVVHVPVYGVKHYTM 605
Qy 541 RNSAEPDQVET--VFASVDPVGNVNGRDEGPGVFIAEVRSR-----NFTNDVAI 590
Db 606 FNOA-----LTQTGRLSSVEPALQNIPIRLERGRKIRQAFVSEBDMILFAADYQIETR 661
Qy 591 RAGVSESVDEETVE-----MSVADMLKE--ASVKILAAGAVALGILSLFSQ 634

```

```

Db      662 LAHIAE---DDNLIEAFRRGLDHTKTAMDIFHVSEEDVTANMRQAQAVNFGLVGIISD 718
      635 KYFLKSSSSFORQDMVSMESDVATIGSVRADSEALPRMDARTAEINIVSKQKISLAF 694
      719 -YGLAQNLTNTRKKAAPFIERFYFASPPGVK-----QYMDNIVQE-AKQKGYVT 764
Qy      695 GPDHRIEMLPEVLADGRM-LKIWTDRAA-ETAOQLGVYDYLTKLSVD-SYTVSADGTRA- 750
      765 TLHRRRYLPDITSRNPNVRSFAERTAMNTPIGSAD-ITKKAMIDLSVRLREERLQAR 823
Db      751 -----LVEATLESACLSDLVHPE 769
      824 LLLQVHDELILEAPKEIERLCRLV-PE 850

```

RESULT 7

```

US-08-436-664-32
; Sequence 32, Application US/08436664
; Patent No. 5874282
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,664
; FILING DATE: 08-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/394,232
; FILING DATE:
; APPLICATION NUMBER: 08/307,410
; FILING DATE: 16-SEP-1994
; APPLICATION NUMBER: 08/222,612
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fisher, Carlos A
; REGISTRATION NUMBER: 36,510
; REFERENCE/DOCKET NUMBER: GP94003.CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-535-2807
; TELEFAX: 619-546-7929
; TELEX:
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-436-664-32

```

```

Query Match      3.28; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0053;

```

```

Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
Qy      99 HPLTGDIRRAPEARVSKPQFGSDALISRRQILQACETLSPRSRRRY-NEGILLDE 157
      107 HYEADDLIGTMAARARE---GFAVKYISGDRULTQ-----LASPQYVETITKKGIIDIE 158
Qy      158 EATVITDV-----PMDKVG-----ALCVLQEGGEIYILRVG 190
      159 SYPTETVEKYGTLPEQIVDLKGLMGDKSDNIGVPGIGKTVKLLKQPGTEVENVLASI 218
Qy      191 BALLKERLPKSPFO--DVVLWALAFLDVSRDA---MADPPPTIGYEEVEALKLLQE 245
      219 DEIKGEKLENLQYRDIAL-LSKOLAICRDPVELTLD--DIVYKGEERKVALFOE 275
Qy      246 EGASSLAPDLRAQIDE-----TLEBITPRYVLBLGI---PLGDDYAKRLNG- 290
      276 LGFQSLDKMAVQTDSEKPLAGMDFAIDSVTDEMADKALVVEVGGNNHAPVIGI 335
Qy      291 -----LSGVNRLKSVGGGASALVGLTREKFNNEAFLRMTAAE--- 330
      336 ALANERGRFFLRBETALADPKFLAW-----LGDETKKKTMDPSKRAVALKWKGI 385
Qy      331 -----QVDLEFVATPSPNIPAES-----PEVVEVALALVNOAFTGK-----KPHLQ 370
      386 ELRGVVFDELALAYLLDPAQAGDVAAVAMQGYEAVRSDEAVYGKAKRTVDEPTLAE 445
Qy      371 DADKQFQLOQAKVMMEIPAMLDTNNWE-IDFGLEGLCALLIG-----KVDECRM 423
      446 HLAR-----KAAAIWMBEPLMDLRNEDRLTELBEQLAGILANMEFTGKVDYKRL 500
Qy      424 -WIGLD-SEDSQYRNPAIYEFVLNSNRDNDLPGLCLELFWLGVVPPRRDTKDK 481
      501 EOMGAEILTEQLQ---AVERRIYELAQEFNINSP--KOLGT---VLPDKQLPVLKK 549
Qy      482 FKLDYVDPMVLSYL-ERREVVQSGSPILAAATMAATIGAEHVASAMQALQKVPSPRYTD 540
      550 TKTG-YTSADVLEKLAIPHHEIVE--HILHYROLQKQSYTIE-GLUKVHPVYTKVHTM 605
Qy      541 RNSAEPDVOET--VPSVDPVGNVGRDGEPPGVFIABAVPSE-----NFEINDYAI 590
      606 FNQA---LGTGRLSVSENLQNIPIRLEEGKIKQAFPSPDMLIFADVSQJELRV 661
Qy      591 RAGVSESSVDETTE-----NSVADMLKE--ASVKIILAAGVALGILISFSQ 634
      662 LAHIAE---DDNLIEAFRRGLDHTKTAMDIFHVSEEDVTANMRQAQAVNFGLVGIISD 718
Qy      635 KYFLKSSSSFORQDMVSMESDVATIGSVRADSEALPRMDARTAEINIVSKQKISLAF 694
      719 -YGLAQNLTNTRKKAAPFIERFYFASPPGVK-----QYMDNIVQE-AKQKGYVT 764
Qy      695 GPDHRIEMLPEVLADGRM-LKIWTDRAA-ETAOQLGVYDYLTKLSVD-SYTVSADGTRA- 750
      765 TLHRRRYLPDITSRNPNVRSFAERTAMNTPIGSAD-ITKKAMIDLSVRLREERLQAR 823
Qy      751 -----LVEATLESACLSDLVHPE 769
      824 LLLQVHDELILEAPKEIERLCRLV-PE 850

```

RESULT 8

```

US-08-436-664-34
; Sequence 34, Application US/08436664
; Patent No. 5874282
; GENERAL INFORMATION:
; APPLICANT: RIGGS, MICHAEL G.
; APPLICANT: SIVARAM, MATHOOR
; APPLICANT: TUDOR, STARLA D.
; TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gen-Probe Incorporated
; STREET: 9880 Campus Point Drive

```

CITY: San Diego
 STATE: CA
 COUNTRY: USA
 ZIP: 92121
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/436,664
 FILING DATE: 08-MAY-1995
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/394,232
 FILING DATE:
 APPLICATION NUMBER: 08/307,410
 FILING DATE: 16-SEP-1994
 APPLICATION NUMBER: 08/222,612
 FILING DATE: 16-SEP-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Fisher, Carlos A
 REGISTRATION NUMBER: 36,510
 REFERENCE/DOCKET NUMBER: GP94003.CP2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-535-2807
 TELEFAX: 619-546-7929
 TELEX:
 INFORMATION FOR SEQ ID NO: 34:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 876 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: internal
 ORIGINAL SOURCE:
 US-08-436-664-34

Query Match 3.2%; Score 128.5; DB 2; Length 876;
 Best Local Similarity 21.0%; Pred. No. 0.0053;
 Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
 QY HFLTDGIRAPFARVSKPQGFSDALISRRQILQAACETLSNPSRRRY-NEGLDDR 157
 DB 107 HYRADDIIGTMAARAERE--GFAVKVISCGRDLTQ--LASPVVETTKKGIIDIE 158
 QY 158 EATVITLV-----PMKVPG-----ALCVQOEGETIVLRVG 190
 DB 159 SYPEYVERKYLTPQIIVDLKGLMGKSDIPGVPGIGEKTAVALKKQFTVENVASI 218
 QY 191 EALKERLPKSPKQ--DVLVLMALAFDVSRA---MALPPDPITGYEFVEELKLQOE 245
 DB 219 DEIKGEIKENLQYRDLAL-LSKQLAICRDAAVELTLD--DIVYGEDEKRYVALFOE 275
 QY 246 EGASSLAADLRADIDR-----TLSEITPRVLELIGL--PLGDDYAAKRING- 290
 DB 276 LGQSFIDKKAAYQVDBGEKPLAGMDFAIDSVTDEMLADKAAVAVRVGNDYHNAPIVGI 335
 QY 291 -----LSGVNIIWSVGGGASALVGGTLREKPMNAFLMNTAAE----- 330
 DB 336 ALANERGFPLRPETALADPKFLAM-----LQDETICKKMPDSKRAAVALKKGI 385
 QY 331 -----QVLDLFAVATPSNIPASS-----FEVYEVALALVAQAFYIK-----KPHLLQ 370
 DB 386 ELKGVVVDLLAAVILADPAQAAGVAVAAKHQYBAVRSDBAVVGKAKRTPDEPTLAE 445
 QY 371 DADKQFOOLQAAKYAMEIPAMLYDTNNNE-IDFGLEKGCALLIG-----KVDECRM 423
 DB 446 HLAH-----KAAAIWALEBPLMDELRRNQDRLLTELEQPLAGILAMNEFTGVAVDTKRL 500

QY 424 -WLGID-SEDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWAGVPRFRDTRDKK 481
 DB 501 EQMGAELTEQLQ-----AVERRIYELAGQEFNINP---KQLGT-----VLPDKQLPVLRK 549
 QY 482 FKGDYDDPMVLSYL-ERVEVVOGSPLAAMTARIGAEHVKASAMQLOKVPSPRTD 540
 DB 550 TKTG-YSTSDAVLEKLPHEHIVE--HILHYRQLGKQSTYIE-GLKVPVPGKVATM 605
 QY 541 RNSAEPDVOET--VPSADPVGNVVGDSRGVFIAAVPSE-----NFEINDVAI 590
 DB 606 FNQA---LTQGLSSVEPQLONIPRLERGRKIRQAFVPSBDMLIPADYQIETLRV 661
 QY 591 RAGVSESSVDETVY-----MSVADLKE---ASVKILAGAVALGHSISFSQ 634
 DB 662 LAHIAE---DDNLIEARRGLDITKTAMDFHVSSEBDYANMRQAKAVFGIVYGISD 718
 QY 635 KYFLKSSSFORKDQVSSMESDVATIGSVRADSEALPRMDARTAEINIVSKQIKSLAF 694
 DB 719 -YGLAQMINITRKKAAEFIERFYFASPPGVK-----QYMDNIVOE-AKQKGYVT 764
 QY 695 GPDRIEMLEPVLDGRM-LKIMTDRAA-ETAGLGLVYDYLTKLSVD-SYVSADGTRA- 750
 DB 765 TLHRRRYLPDITSRNPNVNSFARTAMTPIQSAD-ITKKAMIDLSVRLREBRLOAR 823
 QY 751 -----LYEATLESACLSDLVHPE 769
 DB 824 LLQVHDELLIEAPKEIERLCRLV-PE 850

RESULT 9
 US-09-135-642-20
 Sequence 20. Application US/09135642
 Patent No. 6066483
 GENERAL INFORMATION:
 APPLICANT: RIGGS, MICHAEL G.
 APPLICANT: SIVARAM, MATTHUR
 APPLICANT: TUDOR, STARLA D.
 TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
 TITLE OF INVENTION: STREPTOTERMOPHILUS
 NUMBER OF SEQUENCES: 34
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Gen-Probe Incorporated
 STREET: 9880 Campus Point Drive
 CITY: San Diego
 STATE: CA
 COUNTRY: USA
 ZIP: 92121
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/135,642
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/394,232
 FILING DATE:
 APPLICATION NUMBER: 08/307,410
 FILING DATE: 16-SEP-1994
 APPLICATION NUMBER: 08/222,612
 FILING DATE: 16-SEP-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Fisher, Carlos A
 REGISTRATION NUMBER: 36,510
 REFERENCE/DOCKET NUMBER: GP94003.CP2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-535-2807
 TELEFAX: 619-546-7929
 TELEX:
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:

/ LENGTH: 876 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / HYPOTHETICAL: NO
 / ANTI-SENSE: NO
 / FRAGMENT TYPE: internal
 / ORIGINAL SOURCE:
 / US-09-135-642-20

Query Match 3.2%; Score 128.5; DB 3; Length 876;
 Best Local Similarity 21.0%; Pred. No. 0.0053;
 Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

```

QY 99 HFLTDGIRAFARFARVSKPPQFGSDDALISRQILQAACTLSNPSRRRY-NEGILDE 157
DB 107 HYADDLIGTMAARARE--GFAVKVISGDRDLTQ-----LASPVTVETTKGIDIE 158
QY 158 EATVITDV-----PMDKVPG-----ALCVLQEGGETEIVLRVG 190
DB 159 SYTPETVVEKYGILTPEDIVDLKGLMGDKSDNIPGVGIGKTAVKLLKQGTVENVLASI 218
QY 191 EALLKERLPRKSPQ--DVIWMALAFUDVSRDA--MALDPDFITGYEVEEALKULO 245
DB 219 DEIKGEKLEKENLQYRDIAL-LSKQLAAICRDAFVELTLD--DIVYGEDEKREKVALFOE 275
QY 246 EGASSLAPDLRAQIDE-----TLBEITPRVYLELLG--PLGDDYAKRLNG- 290
DB 276 LGFQSFUDKMAVQTDGEEKPLAGMDPAIADSVTDEMADKAAALVEVVGNDYHHAPYVGI 335
QY 291 -----LSGVNIIILSVGGGASALVGLTREKPMNEAFRLMTAAE---- 330
DB 336 ALANERGRFLRPBTALADKFLAW-----LQDEIKKKTKMFPKSKAAVALKKKGI 385
QY 331 -----QVDFVATPPSNIPAS-----FEVEVALALVAQAFIGK-----KPHLLQ 370
DB 386 ELRGVFDLILAAVYLDPAQAGDVAALVAKMGQEAVERSDAAYGKAKRTVPDEPTLAE 445
QY 371 DADKQFOQLQQAQVMAEIPAMLYDTRNNWE-IDFGELGICALLIG-----KVDCRM 423
DB 446 HILAR-----KAAIIMALEEPLMBELRNEODRLTELEQPLAGILANMEFTGVAVDTKRL 500
QY 424 -WGLD-SEDSQYRNPAIVEFVLNSNRDNDLPGCLLETWLAGVFPFRDTRKOK 481
DB 501 EQMGABETBQLQ---AVERRIYELAGQFNINSP--KQLGT---VLFDKQLPLVKK 549
QY 482 FKLGDIYDDPMVLSYL-ERVEVVOGSPLAATMARI GAHVKASAMQALQKPPSRITD 540
DB 550 TKTG-YSTADVLEKLAIPHHEIVE--HILHYRQLGKQSTYIE-GLIKVHPVVGKVTM 605
QY 541 RNGAEPDVOET--VFSDVPVGNVGRDGEQVIAEAVPSE-----NPFINDYAI 590
DB 606 FNQA---LQGTGRLSVEEPLNLOINIPRLBEGKIRQAFVPSDEDMILPADYSQIELRV 661
QY 591 RAGVSESVDETVE-----MSVADMLKE--ASYKILAGVAILGLISLFSQ 634
DB 662 LAHIAE---DDNILEARRGIDITHTKAMDIPHVSEEDVANMRQAKAVNFGIVGID 718
QY 635 KYFLKSSSSQKQKQWMSMESDVATTIGVRADSEALPRMDARTAEIVISKQKISLAF 694
DB 719 -YGLAQNMLNITRKEALFEIRYFASFPQVK-----QYMDNIVIE-AKQKQYVT 764
QY 695 GPHRTMLPEVLDGRM-LKIMTDRAA-ETRAQLGLVYDYTLKKLSV-SVTSVADGRRA- 750
DB 765 TLHRRRLPDITSRNFNVSPAFERTAMNPIQSSAAD-IITKAMIDLVSRLREBLQAR 823
QY 751 -----LVEATLESACLSDLVHPE 769
DB 824 LLLQVHDELLEAPKEIERLCRLV-PE 850
  
```

RESULT 10

US-09-135-642-32
 / Sequence 32, Application US/09135642
 / Patent No. 6065483
 / GENERAL INFORMATION:
 / APPLICANT: RIGGS, MICHAEL G.
 / APPLICANT: SIVARAM, MATTHEW
 / TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
 / TITLE OF INVENTION: STEAROTHERMOPHILUS
 / NUMBER OF SEQUENCES: 34
 / CORRESPONDENCE ADDRESS:
 / ADDRESSER: Gen-Probe Incorporated
 / STREET: 9880 Campus Point Drive
 / CITY: San Diego
 / STATE: CA
 / COUNTRY: USA
 / ZIP: 92121
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Diskette
 / COMPUTER: IBM Compatible
 / OPERATING SYSTEM: DOS
 / SOFTWARE: FastSeq Version 1.5
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/09/135,642
 / FILING DATE:
 / CLASSIFICATION:
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: 08/394,232
 / FILING DATE:
 / APPLICATION NUMBER: 08/307,410
 / FILING DATE: 16-SEP-1994
 / APPLICATION NUMBER: 08/222,612
 / FILING DATE: 16-SEP-1994
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Fisher, Carlos A.
 / REGISTRATION NUMBER: 36,510
 / REFERENCE/DOCKET NUMBER: GP94003.CP2
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 619-535-2807
 / TELEFAX: 619-546-7929
 / TELEX:
 / INFORMATION FOR SEQ ID NO: 32:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 876 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / HYPOTHETICAL: NO
 / ANTI-SENSE: NO
 / FRAGMENT TYPE: internal
 / ORIGINAL SOURCE:
 / US-09-135-642-32

Query Match 3.2%; Score 128.5; DB 3; Length 876;
 Best Local Similarity 21.0%; Pred. No. 0.0053;
 Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

```

QY 99 HFLTDGIRAFARFARVSKPPQFGSDDALISRQILQAACTLSNPSRRRY-NEGILDE 157
DB 107 HYADDLIGTMAARARE--GFAVKVISGDRDLTQ-----LASPVTVETTKGIDIE 158
QY 158 EATVITDV-----PMDKVPG-----ALCVLQEGGETEIVLRVG 190
DB 159 SYTPETVVEKYGILTPEDIVDLKGLMGDKSDNIPGVGIGKTAVKLLKQGTVENVLASI 218
QY 191 EALLKERLPRKSPQ--DVIWMALAFUDVSRDA--MALDPDFITGYEVEEALKULO 245
DB 219 DEIKGEKLEKENLQYRDIAL-LSKQLAAICRDAFVELTLD--DIVYGEDEKREKVALFOE 275
QY 246 EGASSLAPDLRAQIDE-----TLBEITPRVYLELLG--PLGDDYAKRLNG- 290
DB 276 LGFQSFUDKMAVQTDGEEKPLAGMDPAIADSVTDEMADKAAALVEVVGNDYHHAPYVGI 335
  
```



```

QY 291 -----LSGVNIIIMSVGGGASALVGLTRKPKMNEAFLMTAA----- 330
DB 336 ALANERGFPLRPETALADPKFLAM-----LDDETKKTKMFKPSKRAVALKMKGI 385
QY 331 -----QVDFVATPSNIPAS-----FEYEVALLVAQAFIGK-----KPHLLQ 370
DB 386 ELRGVFDLILAAVLDLPAQAGVAVAAMHQAARSDAVYGKAKRTVPDEPTLAE 445
QY 371 DADKQFOLOQAKVMAEIPAMLYDTNNWE-IDFGLERGLCALLIG-----KVDECRM 423
DB 446 HLAH-----KAAAIWALEBPLMDLRNEDRLTELEQPLAGILAMFETGVAVDTKRL 500
QY 424 -WGLD--SEDSQYRNPAIVEFVLNSNRDNDLPGCLKLETLWLAGVPPRFDTDKX 481
DB 501 EOMGAELTEQLQ-----AVERRIYELAQEFNINSP-----KQIGT-----VLEDKLQPLVKK 549
QY 482 PKLDYDDPMVLSYL-ERVEVVGSPPLAAATMARIAGHVKASAMQALQKPPSRYT 540
DB 550 TKTG-YSTADVLEKLAHPHEIYE--HILHYRQGLQSTYIE-GLKVVHPVGVKHTM 605
QY 541 RNSAEPDVOET--VFSVDVGVNNVGDGPGVFIABVPSF-----NFTNDYAI 590
DB 606 FNOA-----LTQGRLSSEBPLQNIPIRLBGRKIROAFVPSBEDWLIIFADYSQIELRV 661
QY 591 RAGVSESSVDETYE-----MSVADMLKE--ASVYILLAGVAIGLISLFSQ 634
DB 662 LAHIAE--DDNLLEAFRRGLDHTKTAMDI FHVSEBDVTANMRKAKAVNFGVYGISD 718
QY 635 KYFLKSSSFQKDMVSMESDVATIGSVRADSEALPRMDARTAEINIVSKMOKIKSLAF 694
DB 719 -YLAQNMLNTRKEAFAFIERYPASFPQV-----QYMDNIYGE-ANQKQYVT 764
QY 695 GPDHRIEMLEPEVLIDGRM-LKIWTDRAA-ETNAQLGLVYDTYLLKLSVD-STVVSADGTRA 750
DB 765 TLHRRRYLPDITSRNFNVSFAERTAMNTPIQSAD-IIKAMIDLVSRLREERLQAR 823
QY 751 -----LVEATLESACLSDLVHPE 769
DB 824 LLIQVHDELTEAPKEIERLCRLV-PE 850

```

RESULT 11
US-09-135-642-34
Sequence 34, Application US/09135642
Patent No. 6066483
GENERAL INFORMATION:
APPLICANT: RIGGS, MICHAEL G.
APPLICANT: SIVARAM, MATHOOR
APPLICANT: TUDOR, STARLA D.
TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS
TITLE OF INVENTION: STEAROTHERMOPHILUS
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gen-Probe Incorporated
STREET: 9880 Campus Point Drive
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/135,642
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/394,232
FILING DATE:
APPLICATION NUMBER: 08/307,410

```

FILING DATE: 16-SEP-1994
APPLICATION NUMBER: 08/222,612
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fisher, Carlos A
REGISTRATION NUMBER: 36,510
REFERENCE/DOCKET NUMBER: GP94003, CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-2807
TELEFAX: 619-546-7929
TELEX:
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-09-135-642-34

Query Match
Best Local Similarity 21.0%; Pred. No. 0.0053;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;

3.2%; Score 128.5; DB 3; Length 876;

QY 99 HPLTDIRRAFEARVSKPPQPGSDMLISRRQILOACETLSNPSRRRY-NEGILDD 157
DB 107 HYEADIDITGMAAREE--GFAVKVISGDRDLTQ-----LASPVYVEITKKGIDIE 158
QY 158 EATYITDV-----PMDKVRG-----ALCYVQEGSETEIVRVG 190
DB 159 SYTPEYVEYKGLTPEQIVDLKGLMGKSDNIRKVPGEIGETKAKLKGSTYENVALSI 218
QY 191 EALIKERLPSKFKQ--DVVLWMLAFLDVSRA--MALDPPFITGYEFVEBALKLLQE 245
DB 219 DEIKGEKLENLQRYRLAL-LSKQLAIGCRDAVVELTLD--DIVYGEDEKVALFOE 275
QY 246 EGASSLAPDLRAQIDE-----TLBEITPRVYLELGL--PLGDDYAAKRLNG- 290
DB 276 LGFQSFIDKMAVQTDSEBEKPLAGMDFAIADSVTDEMADRAALVVEVVGDNVHHAPYGI 335
QY 291 -----LSGVNIIIMSVGGGASALVGLTRKPKMNEAFLMTAA----- 330
DB 336 ALANERGFPLRPETALADPKFLAM-----LDDETKKTKMFKPSKRAVALKMKGI 385
QY 331 -----QVDFVATPSNIPAS-----FEYEVALLVAQAFIGK-----KPHLLQ 370
DB 386 ELRGVFDLILAAVLDLPAQAGVAVAAMHQAARSDAVYGKAKRTVPDEPTLAE 445
QY 371 DADKQFOLOQAKVMAEIPAMLYDTNNWE-IDFGLERGLCALLIG-----KVDECRM 423
DB 446 HLAH-----KAAAIWALEBPLMDLRNEDRLTELEQPLAGILAMFETGVAVDTKRL 500
QY 424 -WGLD--SEDSQYRNPAIVEFVLNSNRDNDLPGCLKLETLWLAGVPPRFDTDKX 481
DB 501 EOMGAELTEQLQ-----AVERRIYELAQEFNINSP-----KQIGT-----VLEDKLQPLVKK 549
QY 482 PKLDYDDPMVLSYL-ERVEVVGSPPLAAATMARIAGHVKASAMQALQKPPSRYT 540
DB 550 TKTG-YSTADVLEKLAHPHEIYE--HILHYRQGLQSTYIE-GLKVVHPVGVKHTM 605
QY 541 RNSAEPDVOET--VFSVDVGVNNVGDGPGVFIABVPSF-----NFTNDYAI 590
DB 606 FNOA-----LTQGRLSSEBPLQNIPIRLBGRKIROAFVPSBEDWLIIFADYSQIELRV 661
QY 591 RAGVSESSVDETYE-----MSVADMLKE--ASVYILLAGVAIGLISLFSQ 634
DB 662 LAHIAE--DDNLLEAFRRGLDHTKTAMDI FHVSEBDVTANMRKAKAVNFGVYGISD 718
QY 635 KYFLKSSSFQKDMVSMESDVATIGSVRADSEALPRMDARTAEINIVSKMOKIKSLAF 694

```

Db 719 -YGLAQNLTTRKEAEFIRYFASFGVK-----QYMDNIYVE-AKQKGYVT 764
Qy 695 GPDHRIEMLEPVLDGRM-LKIWTDRAA-ETPAQLGLVYDYTLTKLSVD-SYVSADGTRA- 750
Db 765 TLHRRRYLPDITSRNNVNSFAERTAMTPIQSAA-ITKKAMIDLVSRLREERLOAR 823
Qy 751 -----LVEATLEBSACLSDLVHPE 769
Db 824 LLLQVHDELILBAPKEIERLCRLV-PE 850
RESULT 12
US-08-394-232A-20
Sequence 20, Application US/08394232A
Patent No. 6100078
GENERAL INFORMATION:
APPLICANT: RIGGS, MICHAEL G.
APPLICANT: SIVARAM, MATHOOR
APPLICANT: TUDOR, STARLA D.
TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS STEAROTHERMOPHILU
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gen-Probe Incorporated
STREET: 9880 Campus Point Drive
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/394,232A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/307,410
FILING DATE: 16-SEP-1994
APPLICATION NUMBER: 08/222,612
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fisher, Carlos A
REGISTRATION NUMBER: 36,510
REFERENCE/DOCKET NUMBER: GP94003.CP2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-2807
TELEFAX: 619-546-7929
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-394-232A-20
Query Match 3.2%; Score 128.5; DB 3; Length 876;
Best Local Similarity 21.0%; Pred. No. 0.0053;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
Qy 99 HPLTDGIRAFARVSKPQFGSFDALISRQILQACETLSNPRSRREY-NEGLLDDE 157
Db 107 HYAADLIGMABARAENE--GPAVKYISGDRDLTQ-----LASPYVVEITTKGIDIR 158
Qy 158 EATVITDV-----PMDKVPG-----ALCVLQEGGETEIVLRVG 190

Db 159 SYTPEYVERKYGLTPQIYDLKGLMDKSDNIGVGPIGEKTYVKLLKQFGVENVLAI 218
Qy 191 EALIKERLEPKSPKO--DVVLWMLAFIDVSRDA---MALDPDFITGYEVEEALKLQ 245
Db 219 DEIKGEKLEKNLQRYDIAL-LSKQLAAICRDAPEVELTD--DIVYKGEERKVVALLFOE 275
Qy 246 EGASSIAPDLRAQID-----TLEETPRVYLELGI---PLGDDYAAKRLNG- 290
Db 276 LGFQSPFLDKMAVQTEDEGEKPLAGMDPALADSVYDEMILADKAAVLVEVGDYHNAPIVGI 335
Qy 291 -----LSGVRNILNSVGGGAGALVCGLTRKFMNEAFLRMTAAE----- 330
Db 336 ALANERGFPLRPETALADPKFLAW-----LGDEYKKTWEDSKRAVALKMKGI 385
Qy 331 -----QVDFVATPSPNIPAES-----FEYEVALLVAQAFIGK-----KPHLLQ 370
Db 386 ELRGVVPDILLAAVLDPPQADGVAAVAKMGOEAVRSBAVYGAKAKTVEDEPTLAE 445
Qy 371 DADKQFQQLQQAQKVMMEIPAMLYDTRNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db 446 HLAR-----KAAIWMLEBPLMDELRRNEQDRLLTELEQPLAGILANMEFTGVVDTKRL 500
Qy 424 -WIGLD-SEDSQRYNPAIVEFVLENSNRDNDLPLGLCKLFTMTWAGVPRPRDTDKX 481
Db 501 EOMGABLTQQLQ---AVERRIYELAQEFNINSP---KQLGT-----VLFDKLQLPVLRK 549
Qy 482 FKLGDYDDPMVLSYL-ERVEVVQSGSPLAAATMAIRIGAEHVYASAMQALQKVPSPRYTD 540
Db 550 TKTG-YTSDADVLEKLAAPHIYIE--HILHYROLGLQSTYIE-GLLKVHPPTGKHTM 605
Qy 541 RNSAEPDVQET--VPSVDVGNNGVRDGEVGFIAVAPRS-----NEFTNDYAI 590
Db 606 FNOA---LTQTRLSSVEBNLQNIPIRLSEGGKRIQAFVPSBPDWLIFAADYSQIELRV 661
Qy 591 RAGVSESSVDETYE-----MSVAMLKE---ASVKILAAGAILGLISLFSQ 634
Db 662 LAHIAE---DDNILEAPRGLDIHTKTAMDFHVSEBDVTANRRQAKANFGIYGISD 718
Qy 635 KYFLKSSSFORKDVMYSMSDVATIGSVRADSEALPRMDARTAEIVYSKOKISLAF 694
Db 719 -YGLAQNLTTRKEAEFIRYFASFGVK-----QYMDNIYVE-AKQKGYVT 764
Qy 695 GPDHRIEMLEPVLDGRM-LKIWTDRAA-ETPAQLGLVYDYTLTKLSVD-SYVSADGTRA- 750
Db 765 TLHRRRYLPDITSRNNVNSFAERTAMTPIQSAA-ITKKAMIDLVSRLREERLOAR 823
Qy 751 -----LVEATLEBSACLSDLVHPE 769
Db 824 LLLQVHDELILBAPKEIERLCRLV-PE 850
RESULT 13
US-08-394-232A-32
Sequence 32, Application US/08394232A
Patent No. 6100078
GENERAL INFORMATION:
APPLICANT: RIGGS, MICHAEL G.
APPLICANT: SIVARAM, MATHOOR
APPLICANT: TUDOR, STARLA D.
TITLE OF INVENTION: PURIFIED DNA POLYMERASE FROM BACILLUS STEAROTHERMOPHILU
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gen-Probe Incorporated
STREET: 9880 Campus Point Drive
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS

Fri Jun 10 09:57:54 2005

us-10-600-070-2.ra1

Page 13

Job time : 33 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2005, 00:53:22 ; Search time 26 Seconds
(without alignments)
2964.216 Million cell updates/sec

Title: US-10-600-070-2
Perfect score: 4063
Sequence: 1 MEALSHVIGLSPPQLCRLP.....YEVFMWSKSGWKTEGSVLAS 801

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	497	12.2	798	AD2144	hypothetical prote
2	429	10.6	714	S7082	hypothetical prote
3	145	3.6	2297	AB2494	hypothetical prote
4	143	3.5	2089	A48757	acetyl-CoA carboxy
5	139.5	3.4	879	JC4286	DNA-directed DNA P
6	138	3.4	2021	A84771	similar to ch-TOG
7	131.5	3.2	1194	DJBE28	DNA-directed DNA P
8	131.5	3.2	2111	A70668	mycocerosate synth
9	130.5	3.2	1026	C97783	cell surface anti
10	129.5	3.2	2110	B44110	mycocerosate synth
11	129	3.2	1275	T49362	hypothetical prote
12	128.5	3.2	836	T30312	hypothetical prote
13	128.5	3.2	876	S70368	DNA polymerase I -
14	127.5	3.1	2472	E83594	cell frimesshite p
15	126.5	3.1	1018	T30853	antigenic heat-sha
16	125.5	3.1	718	A81122	hypothetical prote
17	125.5	3.1	4385	T29042	hypothetical prote
18	123.5	3.0	673	C63080	probable chemotaxi
19	123.5	3.0	848	C70834	probable endopepti
20	123	3.0	3421	W2BEB6	367k tegument prot
21	122.5	3.0	899	H87513	hypothetical prote
22	122.5	3.0	2756	T30183	hypothetical prote
23	121	3.0	1381	AF2010	regulatory protein
24	120	3.0	1252	D72122	RNA polymerase bet
25	120	3.0	1262	F81548	DNA-directed RNA P
26	120	3.0	3938	T42761	Basoon protein -
27	119.5	2.9	692	G87415	glycyl-tRNA synth
28	119.5	2.9	698	A82593	hypothetical prote
29	119.5	2.9	831	S26675	DNA-directed DNA P

30	118.5	2.9	3942	2	T42730	Basoon protein -
31	117.5	2.9	1004	2	G87323	hypothetical prote
32	117.5	2.9	4613	2	T17409	polyketide synthas
33	116.5	2.9	880	2	B69680	DNA polymerase I p
34	116.5	2.9	1074	2	T01884	hypothetical prote
35	116.5	2.9	1644	2	AC0823	probable lipoprote
36	116.5	2.9	3450	2	T26963	hypothetical prote
37	116.5	2.9	3461	2	T26964	hypothetical prote
38	115.5	2.8	693	2	G82618	plus biogenesis p
39	115.5	2.8	804	2	T44506	transducer protein
40	115	2.8	1983	2	AC1922	two-component hybr
41	114.5	2.8	758	2	A98290	exopolysaccharide
42	114.5	2.8	758	2	A98294	hypothetical prote
43	114.5	2.8	1755	2	F82618	chemotaxis-related
44	114.5	2.8	2333	1	GNNY2F	genome polypolypein
45	114	2.8	777	2	T44597	transducer protein

ALIGNMENTS

RESULT 1

AD2144
hypothetical protein al12707 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. strain PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AD2144
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaoka, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2144
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-798 <KUR>
A:Cross-references: UNIPROT:Q8YTL0; GB:BA000019; PIDN:BA074406.1; PID:g17131800; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al12707
C:Superfamily: Nostoc sp. hypothetical protein al12707, dnaJ amino-terminal homology

Query Match	12.2%	Score 497;	DB 2;	Length 798;
Best Local Similarity	24.6%	Pred. No. 6e-24;		
Matches	207;	Conservative 139;	Mismatches 301;	Indels 196; Gaps 33;
QY	84	VPIFDYOVVGAQTHFLTDGIRRAFEARVSKPQFGFSDALISRRQILQACETLSNP	143	
DB	11	VRIPLDYRILGLPLASDQLRQAYSDRIYQLPRRYSQALISRRQLIEAVVLSDF	70	
QY	144	RSRRRY-----	NEGLDDEAVITVDVPMKYPGALCV	176
DB	71	KERSSYQQLYLAAHYDDNATTKVAVENRGSNNGHFVQSLSI--EVSSEYDILGALLI	128	
QY	177	LOEGGEITVLRVGEALL-----	KERPKSF-----KQDVVLWALAFIDVSR	219
DB	129	LOEIGEYELVKKGRNYLGNONGTASRNGNHRPESFLDSSRPDILLVVALSLSIGR	188	
QY	220	D-----	AMALDPPDFTTGYEFVEBALKLOEBEGASLAPDLRAQIDETLEETPR	269
DB	189	EQWQGHYENALSLF-----	TQGE-----VLFSG--IPFSVQAEIGADLYKLRPY	233
QY	270	VYIELGLPLGDDVYAAKRLGLSGVRNIIWSVG--GASALVGLTBE--KPMNEAFR	325	
DB	234	RIIEELALP--QEKTIERRHQLDLOSLIDRGGIDDTGNDQSLNIDDFLRFQQRHH	291	
QY	326	MTAAEYVDLFAATPSNIPASFEYFEVALVAQAFGKKPHLLQADKQOQOQAKVM	385	
DB	292	LTVAEQKLPDG--ESKRP--AVATYLLAVASINAGFQROKALIRNAKQILMRLSKO--	347	
QY	386	AMEIPAMLYDTRNNWEIDFGIERGLICALLIGKVDCEMMWGLDSEDSQYRNPAIVEFLVE	445	

```

Db 348 -----DVHLEQSLCALLLQGTBEATRVLELSQE---YEALALI---R 383
Qy 446 NSNRDNDLPGICKLETLTAGVFPFRDTPDKKFKLGDYYDDPWLSTYLERV----- 500
Db 384 EKQSDPDLPGICLTAEGQLQNEVFPFRDLSGQASLKDYPANQOVQVLEALPDAE 443
Qy 501 -----EYVQGSPLAAATMAR--IGAEN--VKASMAQLQKVPSP---RYTD 540
Db 444 TTENMAVINRQSPFGPNSGSGTVPAAKRPVGKANPGEASTPVPQRSHSEVNRQFH 503
Qy 541 RNAAEPDQVETVPSPDVPANNVGRDGE-----PGVFIAEAPRSPENFTND 587
Db 504 QNTPEPELETENHRRPSSNFTTARENISTTDAYTDNYPPELPVERASRP----- 555
Qy 588 YAIRAGVS---ESSVDETV-----EMSVADMKEASVKILAA 622
Db 556 --VQPVSGTGTQTPPRQTKRRRRKKRQAVNVRGSHIQOQRSPSTLGRKRLMIVL 613
Qy 623 GVAIGLISLF---SOKY-FLKS---SSSFQRKDWVSMSBDVATIGSVRADSEALPR 673
Db 614 G-SLGGILFMLIVSTTFGWLKXVFPAPSLQGEQLSIQSPPLEIPDKNAQIQSPDVS 672
Qy 674 MDARTANIVSKQKQKLSLAFGPDHRIEMLPVYLDGMLKIYTDRAETQQLGVYDTYL 723
Db 673 LTETETARKIENMLATYASALGAEHKLESINEILITGSALQWMLIALQDRADNRHREYS- 721
Qy 734 IKLVSQSVTVS-ADGTRALVEATLEESACLSDLVH--PENNATDVRTYTRYEVFMSKSG 790
Db 732 HSKVDSISGSDIDPRASGATVRE---LTQPEHQKQKSGSDER-LRRRYELIRDDI 787
Qy 791 WKT 793
Db 788 WRI 790

RESULT 2
S76082
hypothetical protein - Synecchocystis sp. (strain PCC 6803)
C/Species: Synecchocystis sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C/Accession: S76082
R/Kanekeo, T.; Sato, S.; Kocani, H.; Tanaka, A.; Asamiu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
S.
A/Reference number: S74322, MUID:97061201; PMID:8905231
A/Accession: S76082
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-714 <KAN>
A/Cross-references: UNIPROT:Q5555; EMBL:D63999; GB:AB001339; NID:g1001396; PIDN:BA11006
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C/Genetics:
A/Start codon: GTG
C/Superfamily: Nostoc sp. hypothetical protein al12707; dnaJ amino-terminal homology
F:6-70/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 10.6%; Score 429, DB 2; Length 714;
Best Local Similarity 25.1%; Pred. No. 1,2e-19;
Matches 194; Conservative 124; Mismatches 331; Indels 124; Gaps 30;

```

```

Qy 86 IPIDFQVGAQGNFLTDGIRBAFEARVSKPQFGSDALISROILQAACTLSNPRS 145
Db 3 IPIDFRILIGIPQSGGTEIEQAIQDRLQLPRHNSDAATVTLRNQMLAIAYETLRDPEK 62
Qy 146 RREYN---GLNDE--EATVITDVPMDKVP---GALCYLQSGGTEIYLRVGEALLKE 196
Db 63 RQAVDEEMWGMABALGEALPLTPPELSEPGDEIGALLILDLGVELVVKKGEPVLHD 122
Qy 197 RLPKS--FKQDVVLVMAALADLVSRDAMLDPPDFTIGYFVEEA-LKLQEGSGSLAR 253

```

```

Db 123 PNPPAGGLPDQVLLSVILAHMELSRERWQOQ-----YEFATASIKALARIQONDFP 176
Qy 254 DLRAQIDETLEETIPRYVELLGLP-LGDDYAAKRLNGISGVANNILMSVGG-GGASALVG 311
Db 177 ALBAEIRQELRYLRPFRILELLAKEGOGES---QROGGLALLQAMQDRGIGGKGDYS 223
Qy 312 GLTRE---KFMNAFLMTAAEQVDL.FVATPSNI PABSEFVEYVALALVAQAFIGKKPHL 368
Db 234 GLQNDPFLKFIHQRLCHLTVAEQNALFL--PESQPSLVASYLAIVSHLMAGVEKQDPMA 291
Qy 369 LQDADQFOQLQQAQXMAHEIPAMLVDTRNNWIDGLBEGGLALLIGKDECRMLGLD 428
Db 292 IVEAKSLIITQLENCQ-----DLAEKVLITELGGQTEV--VLAALD 330
Qy 429 SEDSQYRNPAIVFVLENSNRDNDLPGICKLETLTAGVFPFRDTPDKKFKLGDYY 488
Db 331 QGD-----PKIVA-GLESKLATGEDPLTAYTTTEQMLSEIYIPYRDLSPFLSPAYF 384
Qy 489 DDPVLSYLERVAVQ-----GSP-LAAATMABIGAEHVKASAM-QALQKVPSPRYT 539
Db 385 NNPVQOQYLEQLBPSFTTNSFASPALISTATESETPMVHSGAALPDRLELTSTVS--- 441
Qy 540 DRNSAEPKQVETVFPVDPGNVGRDGERGVIAEAVRSENFENDVAIRA----- 592
Db 442 -RRGRSPRSRDVFP-----SADNSGLAVT-TLSALIVDTHSLGTNGIGDSTS 491
Qy 593 -GVSESSVDETTYE-MSVADMKEASVK-----ILAGVAILGLISLPSQK----- 635
Db 492 NGFSSNSAPBSTSKHSPRRKKRVTIKPVRFQIFLCLAGIYGVGATALLINNTGDPILG 551
Qy 636 -----YFLKSSSSFORKDWVSMSBDVATIGSVRADSEALPRMDARTANIVSKKQK 688
Db 552 LLEDPLDVFL-----DQPEFIPDEATSNLILISQ---PNFNOQGVQWVQWGLMD 598
Qy 689 IKSILFGPDHRIEMLPVYLDGRLKIYTDRAETQQLGVYDTYLKLSVSTVVS-ADG 747
Db 599 SKGLAFQNDVQALQSVLAPVLAQQRK-AQRDAQKVHYQHEKQLQILAQVNPDP 657
Qy 748 TRALVEATLEESACLSDLVHPENNATDVR-TYTRYEVFMSKSGWKITEGSQL 799
Db 658 NRATVTRARVEISQPTLGNQQKGSATDDLVLRQVLVRHQGVWIKIDQOVV 710

RESULT 3
AB2494
hypothetical protein al17130 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1y
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C/Accession: AB2494
R/Kanekeo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,
N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab
A/Reference number: AB1807, MUID:21595285; PMID:11759840
A/Accession: AB2494
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2297 <KIR>
A/Cross-references: UNIPROT:Q8YU08; GB:BA000020; PIDN:BA878214.1; PID:g17135668; GSPDB:GT
A/Experimental source: strain PCC 7120
C/Genetics:
A/Gene: al17130
A/Genome: plasmid

Query Match 3.6%; Score 145, DB 2; Length 2297;
Best Local Similarity 19.6%; Pred. No. 1.5; 273; Indels 170; Gaps 29;
Matches 136; Conservative 114; Mismatches 114;

```

```

Qy 89 DFYQVGAQGNFLTDGIRBAFEARVSKPQFGSDALISROILQAACTLSNPRSRE 148
Db 273 EFYQIGLAT-----IAHVNEP-----RRREILKALDLVAKTKGSTD 311

```


QY 149 YNE-----GLIDDEEATVITDVPMDKVPGLCVLQEGGETEIVLNGEALLKER 197
| : : : : :
Db 312 FISEVENTSLINLAALDKBERTFVFOALBALPKI-----ENGVN--IFSIIPQLTER 365
| : : : : :
QY 198 LPSFKODVVLVMAFLVDSRDMAIDPP-----DPITGYEVEEALKILOER-----G 247
| : : : : :
Db 366 EPOLLEKTEIVNNLBERGDYGRBALVAIPFKSEPIILQOALEIAVKIANEEYDRKA 425
| : : : : :
QY 248 ASSIAPRLRAQ-----IDETLEBITPRVVLGLPLGDVDAARLNLSGVRNITW 299
| : : : : :
Db 426 LAALAPRLKETQCNVLYLLERIENTWPNFI-----SGVSETLV 465
| : : : : :
QY 300 SVGGGASALVGLTRREKFMNEAFLRMTAAEQVDLFVATPSNIPAE--SFEYVEVALALVA 358
| : : : : :
Db 466 FI-----ASYL--SKSGSKLQKAF-----NIYQNLSEFGYDDALALVA 502
| : : : : :
QY 359 QA-FIGK-KPHLLQDADKQFOOLQO--AKVMA-----EIPAMLYDTRNMEI 402
| : : : : :
Db 503 LAPHLSTLEBPQLQOALKIAINIKEGASKVVALPAVIPHLPQSKQLLEKAFIITQTEY 562
| : : : : :
QY 403 DFGLEBGLCAL-----LIGKVDCEMVLGIDSEDSQYR--NPAT 439
| : : : : :
Db 563 DNARSRLAVVITPHLSDFESHLDKSLBIIEKLVSTYDSARA--QALVAVAPQLKKNPSTL 621
| : : : : :
QY 440 VEFVLE-----NSNRDNDLPGLC-----KLLETWLAGVFPFRDTKDKKFKLGD 486
| : : : : :
Db 622 LQALALEIAKKINSDQCDALMAIAYOLSESEPELLBOYLKIADTYTHSHKRIAIIP 681
| : : : : :
QY 487 YDDPMVLSYLERVEVQSGPLAAATMARIAGEHVASAMQ--ALQKVPSPRYDNRSAE 545
| : : : : :
Db 682 YLRBPQGTETPLQKFKFASIASAACPCL--PAPKSEVLYQALKMLLDANYSDLLRAN 738
| : : : : :
QY 546 PKVOGEVVFSDVPGNNVGDGERGVFAEAVRSENPETMDVAIRAGVSSSDVETTV- 604
| : : : : :
Db 739 ELKTIYVLELKT-----EQKEKVIIEIAKTIK--NDTFKGAELIAPVATHLSEPDQYKIF 790
| : : : : :
QY 605 EMSVADMLK--EASVKILAAVAGVAILGLISLPSQKYLKSSSFORKMWSMESDVAITIG 661
| : : : : :
Db 791 FLEIIEHLKAIENDSSQAKALAVIPIYLSKSSPESLDKAFIAENLQYQSCFPDDLVTILA 850
| : : : : :
QY 662 S-VRADSEALPRMDARTAEINIVSKQIKSLA 693
| : : : : :
Db 851 THLKERECKTLLEQALKEKAKDIDSEYQAOQDFA 883
| : : : : :
RESULT 4
A48757
acetyl-CoA carboxylase (EC 6.4.1.2) - Cyclorella cryptica
C/Species: Cyclorella cryptica
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: A48757
R/Roeseler, P.G.; Ohlrogge, J.B.
J. Biol. Chem. 268, 19254-19259, 1993
A/Title: Cloning and characterization of the gene that encodes acetyl-coenzyme A carboxy
A/Reference number: A48757; MUID:93374903; PMID:8103514
A/Accession: A48757
A/Status: preliminary; nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-2089 <ROE>
A/Cross-references: UNIPROT:Q39478; GB:L20784; NID:G1065903; PIDD:AAA81471.1; PID:G40945
A/Note: authors translated the codon GGC for residue 1834 as Ala
C/Genetic: 25/1; 729/1
A/Intons: 25/1; 729/1
C/Superfamily: human acetyl-CoA carboxylase; biotin carboxylase homology; lipoyl/biotin-
C/Keywords: biotin binding; lipase
P/99-603/Domain: biotin carboxylase homology <BCH>
P/731-603/Domain: lipoyl/biotin-binding homology <LPB>
P/770/Binding site: biotin (lys) (covalent) #status predicted

Query Match 3.5%; Score 143; DB 1; Length 2089;
Best Local Similarity 20.0%; Pred. No. 1.8;
Matches 169; Conservative 113; Mismatches 289; Indels 272; Gaps 44;

QY 73 LPSIDPREBHVPIPIDFYVLAQTHLTDGIRAREAVRSKP-----QGFSDAL 126
| : : : : :
Db 197 LPLNLDK-----LGIFIGPTGVMSVLGPKXIANIILAQYAKYPSIPMSGSFGPPDDGP 250
| : : : : :
QY 127 ISRRQIIQAACETLSNPRSREYNEGLD--DEEATVITDVPMDKVPGLCVLQEG-- 181
| : : : : :
Db 251 LQ-----ADLTEEGTIPMEIFPKGLVTSADAEVIAANKGWEN--GIMIKASEGGCGK 301
| : : : : :
QY 182 -----ETEIYR-----VEALLKERLPKSPQDVVLVMAALFLVDSRDMAIDP 226
| : : : : :
Db 302 GIRFVNDADIRNAFVQVNSEVTSPIFLMQLCKNAHIEVOIYG---DHGNAVALNG 357
| : : : : :
QY 227 PDFTGYEFVEEALKILOEBSGLADPLRAQIDETLEITPRVVLGLPLGDVVAK 286
| : : : : :
Db 358 RDCSTQRRF-----QKIFEEGSPSIVP-----KETHEM-----ELAAQ 391
| : : : : :
QY 287 RLNLGSGVRNILMSVGGASALVGLTRREKFMNEAFLRMTAAEQVDLFVA--TPSNIP 343
| : : : : :
Db 392 RL-----TONIGYQ--GAGTVEYLYNADNKF--FLELNPRLQVHEPVTGIGANLP 441
| : : : : :
QY 344 ABS-----FEVEVALALVAQAFI GKPHLLQDADKQFOOLQOAKVMAEIPA--- 391
| : : : : :
Db 442 ATQOLVAMGILPFIPIRILYGRDAYGTD--IDPLQERYEL--DSHVIABITLENP 498
| : : : : :
QY 392 -----MLYDTRNM--EIDFGLEBGLCALIGKVDCEMVLGIDSEDSQY-- 434
| : : : : :
Db 499 DEGFKPTSGSERIKFPOSTPVMWGYFSVGANGI-----HEPDSQGH 542
| : : : : :
QY 435 ---RNP-----AIVEVLNSNRD--NDDLPGLCLELT-----WLAGVFP 472
| : : : : :
Db 543 LFAKGPAREQARKALVIALKEMEVRGDIRNSVEYLVLLTEAEFKQTITDTSMLDGI-- 600
| : : : : :
QY 473 RFRITKCKEPLGDIYDDPMVLSYLERVEVQSGPLAAATMARIAGEHVASAMQ--- 529
| : : : : :
Db 601 ----KEKSYKV-----EMPSHL-----VVGAIVFAEFHVAVATEEVKESFRKGVS 644
| : : : : :
QY 530 -----LQKVPSPRYDNRSAEKPQOET----- 552
| : : : : :
Db 645 TAGIPGINSFNIEBAYVLDTKPRIVERISPDVYFTLDGNTIDVEVQYTAGALLATFGG 704
| : : : : :
QY 553 ----VFSVD--PVGNNVRDGE---PGVFIAEAVRP--ENFETNDYAIRAGVSES 597
| : : : : :
Db 705 ETRHIFGMDPELGLRLSLDGAIVLMPTIFDPESELRDTVGTGVAVYLDNGATVBAQ--- 760
| : : : : :
QY 598 SVDETYEMSVADMLKASVYKILAAVAGVAILGLISLPSQKYLKSSSFORKMWSME--SD 656
| : : : : :
Db 761 ---QPYVEEAMKMI--MPIKATESG-----KITHNISAGSVISAGDLLASLELKD 806
| : : : : :
QY 657 VATIGSV-----RADSEALPRMDARTAEINIVSKQIKSLIAGPDRHIEMLPEVLDR 710
| : : : : :
Db 807 PSRYKAIETSGKLDINESKYDLEPQKAVNANVL-----GFNLD-----PEVAQAQ 852
| : : : : :
QY 711 MLKIWTDRAAETAGLVYD--YTLKLKLSVDSVTYADGTRALVEATEESACSLDVHPE 769
| : : : : :
Db 853 AIDBATDSAAADLVQVLEBFYAVESQPDGV--IADVVVRILTAYANR-----TLDVVISE 907
| : : : : :
QY 770 NNA 772
| : : : : :
Db 908 NLA 910
| : : : : :
RESULT 5
JC4286
DNA-directed DNA polymerase (EC 2.7.7.7) - Bacillus stearothermophilus
N/Alternate names: DNA polymerase I
C/Species: Bacillus stearothermophilus
C/Date: 14-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 17-Mar-1999
C/Accession: JC4286
R/Phang, S.M.; Teo, C.Y.; Lo, E.; Wong, V.W.T.
Gene 163, 65-68, 1995
A/Title: Cloning and complete sequence of the DNA polymerase-encoding gene (BspPolI) and
A/Reference number: JC4286; MUID:96001245; PMID:7557480
A/Accession: JC4286

A: Molecule type: DNA
A: Residues: 1-879 <PDB>
A: Cross-references: GB:U23149
A: Note: The authors translated the codon GCG for residue 15 as Arg
C: Comment: It has three enzymatic functions: 5'-3' exonuclease activity, DNA synthesis a
C: Genetics:
C: Gene: BstPol I
C: Superfamily: DNA-directed DNA polymerase I
C: Keywords: nucleotidyltransferase

Query Match	3.4%	Score 139.5;	DB 2;	Length 879;
Best Local Similarity	22.0%	Pred. No. 0.76;		
Matches 175;	Conservative 132;	Mismatches 315;	Indels 173;	Gaps 44;

Qy	99	HFLTGISRAFAFARVSKPQFSDALISRQLOAACETISNPSRREY--NEGILADE	157
Db	108	HYEADDITGTHAARARE---GFAVKVISGRDRLTQ-----LASQVVEITKQITITIE	159
Qy	158	EATVITDV-----PMDKVG--ALCVLOEGGETEIVLRVG	190
Db	160	SYTPETVEYKGLTPEQIVLTKGLMGDKSDNIPGVGIGKTAVALTKLQFGTVENVASI	219
Qy	191	EALKERIPKSPKQ--DYVLVMAALAFDVSRA---MALDPDFITGYEVEBALKTLQOE	245
Db	220	DEIGEKXIKENLRQYRDLAL--LSKQLAAICRAPEVLTLD--DIYKGEDREKVALFOE	276
Qy	246	EGASLIADPLRAQIDE-----TLEBITPRVLELIGL-----PGDDYAAARLNG--	290
Db	277	LGFQSFIDKMAVQTDGEKGLAGMPALADSTIDEMLADKALVVEVGDNYHNAPIYGI	336
Qy	291	-LSGVNITLW-----SVGGGASALVGLTREKFM-----NEAFLEMTAAEQV-D	333
Db	337	ALANERGRFRLPETAVALDPKFLAMIGDETKKTKFDSKRAAVALNGKIELIAGVAVDP	396
Qy	334	LFVATPSNIPAES-----FEVEVALATVAQAFTGK-KPHLLDADKQFOOL--QQAK	383
Db	397	LTLAAYLLDPQOAGDVAAVAAMQHYEAVRSDSEAVYGGAKRTVDEPTLAEQVLRKAA	456
Qy	384	VMAMEIPAMLVDTRNNWE-IDFGIERGICALIIG-----KYDECRM--WLGID--SEDSQY	434
Db	457	IMALEEPLMDELRRNEQRLITELHEHALAGILANNEFTGVAVDTRLRLQMGALTEQIO	515
Qy	435	RNPALIVEFLVLENSNRDNDLPGCLKLETWLAGVPPRPFRDQDKKFRGADYDDPMVL	494
Db	516	---AVERRIYELAGQEFNINSP--KOLGT-----VLFPKLQLPVKKTKRTG--YSTSADVL	564
Qy	495	SYL--BRVNVQGSPLAAATMARIGAEHVKASAMQLOKVPFSRYTDBNSAPKOVQST-	552
Db	565	EKLAPHHEIVE--HILHYRQLOKJOSTIIE-GLKLVHAPVYGKATMENOA---LTOTG	617
Qy	553	-VFSYDPAUNNVGRDSEPGVFLAEAVRSE-----NFTENDYAIRAGVSESSVDETT	603
Db	618	RLSSTEPLQNIPIRLBEGRKIRQAFVSEPDWMLFADYISQIELRVLNHLAE---DNL	674
Qy	604	VE-----MSVADMLKE---ASVKTLAAGVALIGLISFQKYFLKSSSSFOFK	647
Db	675	IEAFRRWIDITKTMADIFHVSEEDVTANMRQAKAVNFGIYVIGISD--YGLAQNINIRK	733
Qy	648	DMVSMESDVTATIGSVRADDSBALPRMDARTALENVSKMKOKIKSLAFEPDRHIEMLPEVL	707
Db	734	EAAEFIEIEFASFGVK-----QYMDNIVQD--AKQKGYVTTLLHRRRYLPIT	780
Qy	708	DGRM-LKWTBRAA-ETAQOLGLVYDTLLKLSVD--SVTVSADGTA-----LVEA	754
Db	781	SRNFNVRRFAARTANMTPIQGSAD-IIKKAMIDLSVSVREERLOARLLQGHDBLILEA	839
Qy	755	TLESAGCISDLVHPE	769
Db	840	PKEEIGRGLCRUV--PE	853

C:Similar to chOGG protein from Homo sapiens [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: AB47771
R:Lili, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Xoo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Varadhan, S.E.; Umayam, L.; Taiton, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
N:ature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:1061797
A:Accession: AB47771
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2021 <STO>
A:Cross-references: UNIPROT:Q9ZQN6; GB:AE002093; NID:g4263790; PIDN:AA015450.1; GSPDB:GN
C:Genetics:
A:Gene: At2g35630
A:Map position: 2

Query Match	3.4%;	Score 138;	DB 2;	length 2021;
Best Local Similarity	20.2%;	Pred. No. 3.6;		
Matches 179;	Conservative 122;	Mismatches 315;	Indels 272;	Gaps 43;

```

OY 39 AKXADRLSDNFNFSDSSSSFPATTTATVSLSPSDRPERHVPIDPFYVLQAQT 98
Db 297 ATKMSERBEAAVELTKLSTKKIAPDGFSEICRLKKLI--TVNLNVAVAEAIQAIG-- 352
OY 99 HFLTDGIRRAFEARVSKPQGFSDDALISRRQIILQAACETLSNPSRREYNEGLDDEE 158
Db 353 --LACGLRTHSA-----SSRMLVLLEKL-----KEKGSVTBPLT 388
OY 159 ATVTIDVMDKVPALCVLQEGGETEIVLRVGEALLKERLPKSKODVYLMA-----LA 213
Db 389 QTLQT-----MYKAG-----CLNLDVIGKVKVTVKPKVPLVRSSTLTWLT 431
OY 214 F-LDVRDAMALDPEDFTTGYEFVEEALKLQEBGASSIAPDLR-----AQID----- 260
Db 432 FCLFETSNALLIK-----AKREYVPLCMGLNDG-----TPDVRDAAFSALAAIAXSGVM 481
OY 261 -----ETLEETIRVYLELLGLPLGDDVAAKRLNGLSGRNIIWSVGSGASALVGL 313
Db 482 RPLERSLKLDVDRKKLSLSEMIAGSGGDOA-----GTSSV-TVQSSVGSTATGCLNRY 535
OY 314 TREKEM--NBAFLMTAAE-----QVDLFVAT-----PSNI 342
Db 536 TTSKFMQNSDASFVKKASASMLSGRRPAPSAQSKVGTGKPGGKKDGSVVRNEGSSVSE 595
OY 343 PAESEFEVEVALATAQAQAFIGK--KPHLLQDAKCF--QQLQAQVWAMEI PAMLYPTRN 398
Db 596 PPEDEPAPMGIELEENR-LGSLVKEPETSQLKSSVMKELENTALLKEIEGL----- 648
OY 399 NWEIDFGLE--RGLCALLIGKVDCECRMWLGLDSEDSQYRNPAIVEVLEN--SNRDD- 451
Db 649 -QEIDKSYEILVRLICAV-----PGNNKVVOL-----FSFENYFPSSNLPDF 689
OY 452 -----NDLPLGLCKLETWLAGVFPREPRDKKFKGIDYDDPMVL 494
Db 690 SINSLGTSERVAADITRASAMKCLTAFCEAVGPFVEERL-----FKIMKEHKMKVYL 742
OY 495 S-----YLERVEVQGSPLAAMATMARIGAEH----- 521
Db 743 SEGGLMMVSAVDDPFGVSLKLKDLIDFCKDVGQSSSTAATRNATIKLGLALHKRVGPDIK 802
OY 522 -----VKASAMQALOKVPFSRRTYDNRNSAEPKVOQEIVFSGYDPGNVNVGRGEPGVFIATAE 576
Db 803 GFLNDVKKFALLSALDTEY-EKNPFEQTAAPKRVKT--SVSTSTSGGLSLPREDISTK 859
OY 577 VRPS--EKPEFNDVYAIRAGVSESSVDETTVMVAMMLKEASVKIILAGVAIGLISPSQ 634
Db 860 ITPMLIKFESEDMKRL-----ESTE-----AVNKILEANKRIOPTGIG-----ELFCG 905
OY 635 KYFLKSSSSFKQKDMVSSMESDPAITGSYRADSEALPRMDARTAEINIV-----SKWQK 688

```

```

Db 906 ---LRGRLDSNKNLVM---QTLTTIGGVAAMGPAVEKASKGILLSVLCIKLGNCKMKR 959
Qy 689 IKSLA-----FGPDRIEMLPEVL-----DGRM-----LKIMTDRAETAQGLVYDYT 732
Db 960 ECTLAALDLMTGAVHLDMMPIYIIITLTDGKMGABGRKDLFDML-----TKQLTGLSDPV 1014
Qy 733 ----LKLTVSDSVTVSADGTRALVEATLEBSACLSIDLVPENNAVDVR 776
Db 1015 DAHTLLPASTAMTODKSADYKAAEGCISELIRVSGGEMIKNLKDIO 1062

RESULT 7
Db 7328
CNA-directed DNA polymerase (EC 2.7.7.7) - human herpesvirus 3
C/Species: human herpesvirus 3, varicella-zoster virus
C/Date: 30-Sep-1998 #sequence_revision 30-Sep-1998 #text_change 09-Jul-2004
C/Accession: B27214
R/Davison, A.J.; Scott, J.E.
J. Gen. Virol. 67, 1759-1816, 1986
A/Title: The complete DNA sequence of varicella-zoster virus.
A/Reference number: A27345; MUID:86306657; PMID:3018124
A/Accession: B27214
A/Molecule type: DNA
A/Residues: 1-1194 <DAV>
A/Cross-references: UNIPROT:P09252; EMBL:X04370; NID:g59989; PIDN:CAA27911.1; PID:g60017
C/Genetics:
A/Suprafamily: DNA polymerase
C/Keywords: DNA binding; DNA replication; nucleic acid transferase

Query Match 3.2%; Score 131.5; DB 1; Length 1194;
Best Local Similarity 19.0%; Pred. No. 4.1;
Matches 166; Conservative 122; Mismatches 297; Indels 297; Gaps 44;

Qy 48 SDNFNTSDSS---SSPATATTATLVSLPSIDRPERHV--PIPIFYQLVGAQ---TH 99
Db 205 NDATLNDKNAFHGTSFSPASPSFSEFVEV---ERTDYVYDTQPCAFYRVSPSSKFTN 261
Qy 100 FLVDGIR---RAFAARVSKPPQFGSDALIS-----RQI----- 132
Db 262 YLCDNFHPELCKEGRDATTATRFMDNPGFVSGWYQLKPGVDGERVVRPASHQLTSLD 321
Qy 133 LQAAAC-----ETLSNPSRREYNEGLDDB-----BATVITDVWMDKVPALC 175
Db 322 VEIDCSNDNLQALPNDSSWPDYKLLCFDICKSGSNBLAPDATHLEDL-----VIGISC 377
Qy 176 VL-----QEGGETEIVLRVGEALLKRLPKSEFKQ-----DVLVMAALFDVSRD 220
Db 378 LTVSIPROSLHEHLLFSLGSCDLPQRYVQEMKAGLPEPVLRFDPSEPELLIAFMVLVKQ 437
Qy 221 AMALDPPDFTTYGFVEALKLQEBGASLAPDLRAQIDETLEITPRYVLELGLPLG 280
Db 438 Y---AEFAATGVYIVFDMAFI-----MEKLSISYLSLK----- 468
Qy 281 DDVAARKLNGISGRNLTMSVGGGASALVGLTRBKPMNEAFRMTAAEQDVFVATPS 340
Db 469 DGGGSIIRGLFKI---WDVKGSG-----FQRSKVKINGLSLDMATATE 512
Qy 341 NIPASEFVEYVA-----VALENSNRDNDNDLPGCKLLETLAGV--VEPR 473
Db 513 KTKLSYKLDSSVABEALNBSKRDLPYKIDPGYVASGPNTRGIIIGEYCIDSALVGLKFPK 572
Qy 364 KKHLLDADKQFQOLQAKYMAEIRPMLYDTRNNWEIDGLERGLCALLIGVDECRM 423
Db 573 YLPHELSA-----VARLARITLTKAYD-----GOQVRIYTCILGLASS--- 612
Qy 424 WLGLDSDSOYRNPAYEF-----VLENSNRDNDNDLPGCKLLETLAGV--VEPR 473
Db 613 -RGITLDPGGY--PATFEYKQVIDVGDVBEKMBDESVSPYSGSRRNVYKQARVPD- 668
Qy 474 FRDTKDKKFKLGDYDDPMVLSYLERV--EVVQSGPLAAMATMAIRGAHNVASAMQALQK 532
Db 669 -PDI-----GFYIDPVVVLDFASLVPSIIQAHNL--CFTTLTLNFEYV-----R 710

```

```

Qy 533 VFPERYTDRNSABPKOVETVFSYDP-----VGNNGRDEGPVFIAE--AVRESENET 585
Db 711 LNPEDY-----ATFTYVGKRLFFVRBNV--RESLITGLMDLMLMRK----- 750
Qy 586 NDVAIRAGVSESSVDEFTVENSVDMLKEASVKILA-----AGVAIGLIS----- 630
Db 751 ---AIRARIPSSSDE-----AVLLDKQQAIRKVCNSVYFTVAGQFELCLVVAATVT 802
Qy 631 ----LFGQKYFLKSSSSGQR-----KQVVSMEGDVA--TIGSVADDEALPRMDAR 677
Db 803 TIGRQMLSTRDYTHNNWAAFERPITAPPDISSVLSQKAVEYKVIYGDTSVFIREKGV 862
Qy 678 TAENIVSKMOK-----IKSLAFGPDRIEMLPEVLDGRMLKIMTDRAAETAQGLVYDYL 733
Db 863 SVEGIATIGEMMAHITSLFPCPIKICEKTFI--KLILL-----TKKKYIGVYGGKV 915
Qy 734 LKLSVDSVTVS---ADGTRALVEATL-----ESACLS 763
Db 916 LMKGVDLVRKNNCOFINDYARKLVELLYDDTVSRRAAASACVS 959

RESULT 8
A70668
Mycobacterium tuberculosis
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C/Accession: A70668
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellern, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: A70668
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-2111 <COL>
A/Cross-references: UNIPROT:P96291; GB:Z83858; GB:AL123456; NID:g3261675; PIDN:CAB06108.1
A/Experimental source: strain H37Rv
C/Genetics:
A/Suprafamily: mycobacteric acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hom
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C/Keywords: acyltransferase; carrier protein; coenzyme A; phosphotransferase
F/27-426/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F/534-815/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F/1444-1733/Domain: long-chain alcohol dehydrogenase homology <LADH>
F/1765-1945/Domain: short-chain alcohol dehydrogenase homology <SADH>
F/2016-2096/Domain: acyl carrier protein homology <ACPI>

Query Match 3.2%; Score 131.5; DB 2; Length 2111;
Best Local Similarity 19.9%; Pred. No. 10;
Matches 151; Conservative 82; Mismatches 237; Indels 289; Gaps 33;

Qy 151 EGLIDSEATYITDVP---WDK-----VPALCYLQEGGETEIVLRVGEALLKRL 198
Db 29 ESLRGGD--LVTEIPDRDADDDYDPEPGVPR--SVSRGSG----- 68
Qy 199 PKSPKQDVVLMALAFDVSRDAMALDPPDFI---TGFEFVEALKLLQEBGASLAPDL 255
Db 69 ---FLDDVAGFDAEFQISREASTIPQGRLLLETWEALIEH----- 109
Qy 256 RAQIDETLEITPRYVLELGLPLGDDYAAKRLNGLSGVNRILMSVGGGASALVGLTR 315
Db 110 -----GLDPAISL-----AGSTAVFTGLTH 129
Qy 316 EKPMNEAFPLMTAAEQVDLVATPSNI-----PASEFVEY----- 350
Db 130 EDYL---VLTTTGAGLASPYVVTGLNNNSVASGRIATHTLGLHGPMTPTDTCSSGLMAVHL 186
Qy 351 -----EVALALVQAQFICKPHLL-----ODADKQFQOLQAK 383

```

```

Db 187 ACSRLHDEADLLAGCAVLLBPHASVAASQMLSTGRCHSPDADAGFVNSGECAM 246
      ||| : |||
      : :
QY 384 VMAMEIIPAMLYDRNNNEIDFGLEGLCALLIGKVECRMWLGDSDSQYRNPAIVEFV 443
      : :
Db 247 VLKRLPDLARDGR-----IFAVRGTRATNODGRTER-----LTPMSDAQV---AVRRA 295
      : :
QY 444 LENSNRDDNDLPGCLKLETWLAGVFPFRPRDTKDKFKLGDYDDPMLSYLRYEVV 503
      : :
Db 296 LAAG-----VQBEIVGV-----BAHGTRTPIGPIEYRSIARV-YG 332
      : :
QY 504 QGSPPLAAATMARI GAHVYASAMQALQKV-----PSRYDRMSAPKQVQEVFVS 555
      : :
Db 333 AGTPCALGSAKSNMGHSTASAGTVGLKALISLRHGVPPPLHFNRLPDELSDV----- 386
      : :
QY 556 VDPVGNVVGDEGPGVIAEAVRSENFETNDVA-IRAGVSESVDETTEVMSVADMLKE 614
      : :
Db 387 -----ETGLFVPQAVTWPN--GNDHTPKVAVSSFGMSGTNHAYVEAPAE 432
      : :
QY 615 ASVKTLAAGVA-IGLISLFSQKPYLKSSSFQKDMVSSMESDVATTGSVRADSEALPR 673
      : :
Db 433 ASAPESPFGDAVY-----PRLFMLSSTS-----SDAL-R 461
      : :
QY 674 MDARTENITYSKQK-----IKSLAFGPDHR-----IEMLPYLDGRML 712
      : :
Db 462 QTRAROLATWVEEHQDCVAAADLATLARGRAHREVRTAVVAANLPBLVEGLREVAADAL 521
      : :
QY 713 -----KIW--TDRAETQALGL-----VYDTYLLKLVSDSVTVSVADGTRALV 752
      : :
Db 522 YDAVAGHGDRGPPVWVSGGSSQWAAKGTQLLASEPVAATIAKL---EPVIAASGFSVT 578
      : :
QY 753 EA-TLESACLSDLVHPENNATDVRYTYTREVFWMSKG 790
      : :
Db 579 EAITAQOTVIGIDKVPVAVPAVQVLAATMEQTGVAVPG 617
      : :

```

RESULT 9

```

C97783
cell surface antigen [Imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: C97783
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ri-
science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: C97783
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1026 <KUR>
A:Cross-references: GB:AE006914; PIDN:AL03205.1; PID:G15619756; GSPDB:GN00173
C:Genetics:
A:Gene: sca4

```

```

Query Match 3.2%; Score 130.5; DB 2; Length 1026;
Best Local Similarity 18.8%; Pred. No. 3.7;
Matches 124; Conservative 109; Mismatches 277; Indels 151; Gaps 25;

QY 115 KPQPFESDALLSRQIIQACETLSNPSRREYVNGSLDDEBATYITVPMKV----- 170
      ||| : |||
Db 378 EPPQYKQVPPPIPTNQPLP--ETSQMPOS-QQVNNPLN--TATMLSGSMQDLNLYVN 432
      : :
QY 171 PGALCVLQEGGETEIVLRVGEALLKERTPKSFKQDVVLVMA--LAFDVSRDAMALDPPD 228
      : :
Db 433 AGITKAIDSXKQDILKEATATILNN-----KSDIAEKQNIITALENTVNNKQLKPA 487
      : :
QY 229 FITGYEFEBALKLQEGGASSLAPDLRAOI-----DETLEETPRYVEL--LGLPLG 280
      : :
Db 488 KVTGVAVAVETIKNDQNTPNLERSKMLEATVAIVLNSENLEPKQKQMLEKAVDVGSLK 547
      : :
QY 281 DQVAAKRLNGSLGVRNITLMSVGGGASALVGLTRFENMEAPLRMAAQQVULFVAVTPS 340
      : :
Db 548 DD--ASRAAADIQIDVY-----IKSNLSPEDKMLAVGD 580
      : :

```

```

QY 341 NIPAESFEYVVALAVQAQFIGKPHLLQDADKQFOQLQAKVMA-----MEIPAM 392
      : :
Db 581 KVVVSELSNAEKQKLGLSGVLKKGVEAQVLSPPAQOQLMQHLYYIMAEQTKDITIKVNDI 640
      : :
QY 393 LYDTRNNW-----IDFGLERGLCALLIGKVECRMWLGDSDSQYRNPAI 439
      : :
Db 641 LFDPLSTBELKTNITQAITSNVLDGPATVAEKKELLQATNTTLAGSSLEAQD---KAAI 696
      : :
QY 440 VEPVLN--SNRDNDLPGCLKLETWLAGVFPFRPRDTKDKFKLGDYDDPMLSYL 497
      : :
Db 697 IKVGERTIAHSDTSLSPKALIMASAEKGI-----EGQTNLPD--RELMTGLV 746
      : :
QY 498 ERVEVVGSPPLAAATMARI GAHVYASAMQALQKV-----PSRYTRN 542
      : :
Db 747 DGIYEGKGGEITKAVSSGIDNSININDESEKALKKADAASBALDRDTQNLTEGFKGN 806
      : :
QY 543 SAPPK-----DVQEVFSPVDPVGNVNGDGPVFIAP-----AVR 579
      : :
Db 807 IEERKPHDDIYNKAREVINAVNPVLEALKSKPPVSAERIYQETSSINNISKLAVEK 866
      : :
QY 580 SENFETNDVAIRAGVSESVDETTEVMSVADMLKEASVKIILAGVAILSLFSQKPYLK 639
      : :
Db 867 VNNP-----RMLSPNGMLKTLERK-----KEAIIKRYDE-----LYKAGTGSSTE 908
      : :
QY 640 SSSSFQKDMVS-----SMESDVATTGSVRAD--SEALPRMDARTAE-NIVSRKQIKSL 692
      : :
Db 909 EQGSFIKTNLIDDKTSLKSVRLQTDILQEQKRSKAIENPSVTEBDRVAVSGSKLKPI 968
      : :
QY 693 A 693
      : :
Db 969 S 969
      : :

```

RESULT 10

```

B44110
mycocerotate synthase (EC 2.3.1.111) - Mycobacterium bovis
C:Species: Mycobacterium bovis
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 09-Jul-2004
C:Accession: B44110
R:Mathur, M.; Kolatukudy, P.E.
J. Biol. Chem. 267, 19386-19395, 1992
A:Title: Molecular cloning and sequencing of the gene for mycocerotate acid synthase, a nc
-Guerlin.
A:Reference number: A44110; MUID:92406887; PMID:1527058
A:Accession: B44110
A:Molecule type: DNA; protein
A:Residues: 1-2110 <MAT>
A:Cross-references: UNIPROT:Q02251; GB:M95808; NID:g149977; PIDN:AAA25369.1; PID:g149980
A:Note: Sequence extracted from NCBI backbone (NCBIN:113897, NCBI:P:113899)
A:Note: the source is designated as Mycobacterium tuberculosis var. bovis Bacillus Calmet
C:Superfamily: mycocerotate acid synthase; 3-oxoacyl-[acyl-carrier-protein] synthase I hom
nase homology; [acyl-carrier-protein] S-malonyltransferase homology
C:Keywords: acyltransferase; carrier protein; coenzyme A; phosphotransferase
F:27-426/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
F:534-815/Domain: [acyl-carrier-protein] S-malonyltransferase homology <AMT>
F:1443-1732/Domain: long-chain alcohol dehydrogenase homology <LADH>
F:1764-1944/Domain: short-chain alcohol dehydrogenase homology <SADH>
F:2015-2095/Domain: acyl carrier protein homology <ACP1>

```

```

Query Match 3.2%; Score 129.5; DB 2; Length 2110;
Best Local Similarity 19.3%; Pred. No. 13;
Matches 151; Conservative 82; Mismatches 237; Indels 289; Gaps 33;

QY 151 EGLLDDEBATVITDVP---WDK-----VPGALCVLQEGGETEIVLRVGEALLKERTL 198
      ||| : |||
Db 29 ESLRGDD--LVTEIPDRKRDADYIDPEGVGR-SVSMWG----- 68
      : :
QY 199 PKSFQDVVLVMA-LAFDVSRDAMALDPPDFI---TGYEFEBALKLQEGGASSLAPDL 255
      : :
Db 69 ---FLDDVAGFDEAFPGISREARATSIDPQORLLLETSWEAIEHA----- 109
      : :
QY 256 RAQIDETLEETPRYVELLGLPLGDDYVAAKRLNGSLGVRNITLMSVGGGASALVGLTR 315
      : :

```

Db 110 -----GLDPASL-----AGSSTAVFTGLTH 129
 Qy 316 EKPMNEAFRLMTAAEQVDLFPVATPSNI-----PASEPVY----- 350
 Db 130 EDYL---VLTTTAGLASPYVVTGLPNNSVASGRIAHITGLHGPMTPTTACSSGLMAVHL 186
 Qy 351 -----EVALVAQAFIGKKPHL-----QPADKQFOOLQOAK 383
 Db 187 AGRSLHGERDLALAGCAVALLERHACVAAQAQCMSTGRCHSDADADAFVBSGCAAM 246
 Qy 384 VMAMEIPAMLYDTNNWEIDFGLERGLCALLIGKVDCEKMWIIGLSEDSQYRNPAIVEFV 443
 Db 247 VILKRLFDALRDGR-----IFAVRGATNQDGRRET-----LTPMSDAQV---AVYRAA 295
 Qy 444 LENSRRDNDLDFGLCKLETLWLAGVFPFRDRTDKKFKLGDYDDPMLSTYERREVV 503
 Db 296 LAAG-----VQPEYGVV-----EAMGTGTFGPIEYRSIARV-YG 332
 Qy 504 QGSPLAAATMARI GAHVHVKASAMQALOKVF-----PSRYTDRNSAEKDVQETVFS 555
 Db 333 AGTCALGSAKSNHSTAGTGLIATLISHGVVPLHLFRRLPDELSDV----- 386
 Qy 556 VDPVGNVGRDGEFVIAEAVRPSSEFETNDYA-IRAGVSESSVDETTEMVSADMLKE 614
 Db 387 -----ETGLFVPQAVTTPMN--GMDHTPKRYAVASSFGMSGTNVAHIVEAPAE 432
 Qy 615 ASVKILAGVA-IGLISLFSOKYFLKSSSSFORKDMVSSMESDVATIGSVRADSEALPR 673
 Db 433 ASAPESPFGDAEVO-----PRLFMLSSTS-----SDAL-R 461
 Qy 674 MDARTANIVSKWK-----IKSLAFGPDHR-----IEMLEPVLDGRL 712
 Db 462 QTRAQALATWEEHODCCAASDLATLARGRHRPYRAVAANPELVESGIREVADGAL 521
 Qy 713 -----KIW--TDRAETALGL-----VYDTLLKLSVDSVTVSADGTALV 752
 Db 522 YDAVGHGDRGPVVFSGQSGQMAAMGTQLLASPEVFAATIAKL---EPVIAASGFSVT 578
 Qy 753 EA-TLESACISDLVHPENNATDVRTTYTRYEVMSKSG 790
 Db 579 EAITAQQTVGIDKVPVAVFAVQAALATWETQYGVPRG 617

RESULT 11
 T49362
 hypothetical protein Bld1.160 [imported] - Neurospora crassa
 C:Species: Neurospora crassa
 C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
 C:Accession: T49362
 R:Schulte, U.; Allyn, V.; Hobeisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000
 A:Reference number: Z25022
 A:Accession: T49362
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1275 <SCH>
 A:Cross-references: UNIPROT:Q9P6A4; EMBL:AL355927; GSPDB:GN00116; NCSP:Bld1.160
 A:Experimental source: BAC clone Bld1; strain OR74A
 C:Genetics:
 A:Gene: NCSP:Bld1.160
 A:Map position: 6
 A:Introns: 24/3

Query Match 3.2%; Score 129; DB 2; Length 1275;
 Best Local Similarity 20.9%; Pred. No. 6.5;
 Matches 114; Conservative 84; Mismatches 210; Indels 138; Gaps 26;

Qy 325 RMTAAEQVDLFPVATPSNIPASFEVYVALLVAQAFIGKKPHL----- 368
 Db 39 RMTFATPTAVFAALLSLNNENLATPVVYLLDVVTFP-APRPILIQAKFTQITITLAPLVS 97
 Qy 369 LGDADKQFOOLQOAKVMAEIPAMLYDTNNWEI---DFGLERGLCALLIGKVDCEKMWL 425

Db 98 LPDAD---APLMRASISGLE-SLLADATQWEIGTAQIGRRVAVAGLLA-----L 144
 Qy 426 GLDESDS-QYRNPAIVEFVLENSNRDNDLDFGLCKLETLWLAGVFPFRDTKD----- 479
 Db 145 ALDRPVPKRAQBALNRLINPPSPSLDHPAPMCAETAMQSL-----RLVADKAAQ 199
 Qy 480 -KFKLGDYDDPMLSTYERREVVQ-----GSPLAATAW-----ARIGAEHVKSAM 527
 Db 200 RKEKTTDSTDPRLIALQLVKAVASGGWPKSKLSLCELLSTARTNEHMSAVF 259
 Qy 528 QALQKVPESRYDR-NSAEPKDVQETVFSVDPVGN-----VGRDGEFVIAEA 576
 Db 260 EIFEMFEGMAADSVASAKPLRLLEIRLPAFNDQLLEPWTALISRAVD-----VAGQ 315
 Qy 577 VRSENF--TNDYAIAGVSES-----SVDETTEMVSADMLKERSV-----KIL 620
 Db 316 ISPAETFOELVDPPTLVAGYLSQHKNIKRVASASCLVPLANCVKKEALIEPSIFDEKVI 375
 Qy 621 AAGVAI---GLISLFSOKYFLKS-----SSSFORKDMVSSMESDVATIGSVRADSE 669
 Db 376 QQLKVVYEGTLTVQQAAMETFNVLGAMPAPRWQANPYLLSV---VKSIGELRGDSF 432
 Qy 670 ALPMDARTANIVSKWKIKSLAFGPDHRIEMLP-----EVLDRMLKIWDRAETAO 725
 Db 433 A---GQGEDEVLGKAIK---AMGPBVALNVLPLMLAKVKGQPGRAW-----ML 476
 Qy 726 GLVDTY---LKLSDSVTSVSGDTRALVEATLESACISDLVHPENNATDVRTYTR 781
 Db 477 PLRDLSTNNLHAFKSELVPLSAIMFORVIE-----HOGKNIHIIKIEFTV 523
 Qy 782 YEYFWS 787
 Db 524 VQQTWS 529

RESULT 12
 T30312
 pilin biosynthetic protein - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 04-Mar-2000
 C:Accession: T30312
 R:Whitchurch, C.B.; Young, M.D.; Hobbs, M.; Matick, J.S. submitted to the EMBL Data Library, November 1996
 A:Description: Pseudomonas aeruginosa chemotactic transduction genes pilL, cnpA cnpB and
 A:Reference number: Z20819
 A:Accession: T30312
 A:Molecule type: DNA
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Cross-references: EMBL:U79580; NID:g3241967; PID:g3241968; PIDN:AAC23930.1
 C:Genetics:
 A:Gene: pilL
 A:Superfamily: Pseudomonas aeruginosa pilin biosynthetic protein

Query Match 3.2%; Score 128.5; DB 2; Length 836;
 Best Local Similarity 20.0%; Pred. No. 3.6;
 Matches 158; Conservative 106; Mismatches 292; Indels 235; Gaps 35;

Qy 124 DALISRQIIQAACETSNPRSRREYNEGLDDEATVITDVPMDKYGALCVLQEGGE- 182
 Db 24 ETLKQARQALEAFENQDPTPRK-----FCLTVV--HVGQTLQWMEFYGAA 69
 Qy 183 --TEIVLRVGBALIKERLPKSPKODVVLVMAFLVSRDAMALDPDF-----ITGYE 234
 Db 70 LLAEMEQALQVALLDGVVPGNEBALBVLMOAIIQLPYLDRICFARBDLPVVVPLINDL 129
 Qy 235 FVEBALTLQDEGSSLAPDL---RAQID-----ETLEETPRVYVELGLP 278
 Db 130 RAARGELKLSB--TSLAPDLSQRQPDLDGEALQLRTDELGGILKRLRQTCOMALVGL 187
 Qy 279 LGDDYAAK-----RLNGLSGVRNI--LWSVGGGASALVGLTRERKMEAFPLMT 327

```

Db      188 RNDVATSLGLARVYARLEGLCREAPGLMSI-----ASGLVEGLANGSVNSASVR-T 242
Qy      328 AAEVD-----LFTATPSNI-----PAESFEVEVALVAQAFIG 363
Db      243 LIRQJDEBLKLYEQGADGLQAAPDELVKLLFYAKAPQSFRIRAL----- 291
Qy      364 KKPRLQDA--DKQFOQLQAQVMAWEIPIAM-----LYDTRNWEIDFGLERG-- 409
Db      292 KEQYRLDEALPDHETVAERARLAGPRDANRSVVALCEELVRIKQSLDFVNSDGHF 351
Qy      410 --LCALL--IGKUDCEKMLGLDSEDSQYRNPAIYEPFLSNSDDND---DLPGICL 461
Db      352 SEDLLALAPKQJLDLTAIVGFGQPRKVTLDQDLVDVHIALAQGRREPSDAILMDVAGLLY 411
Qy      462 LETVLAVGVP--RPDTRDKKKFKLGDYD-----DPMVLSYL----- 497
Db      412 VEATLAGMAGBGRNBERSESLPTTDDVAQHVLVYKERRNGLEQAKAIIFIASQNNHE 471
Qy      498 -----ERVEVQGS-----PLAAATMARIGAEHVKASAMQALQKVP----- 535
Db      472 HLAIVPELLTQVRGLAMIFLERAAITLLETGNRYIQEQL--ARKAVPDMQSLDTLADI 529
Qy      536 ---SRYDRNSAEKDVQETFSV--DPVGN-----NGRDEPGVFIAEAVRSENF 583
Db      530 TSEVYLERLSBEDHASQSDLLIDVAEDSLANIGYTLKPNSSAPAPPGI----- 577
Qy      584 ETNDYAIRAGVSESSVDETTVMGVADMLKEASVKIIAAGVAILISLFSQKFLKSSS 643
Db      578 -SGPALESAPAEERPEAVE--VAETAEPDPADTPALEA----- 616
Qy      644 FORKDWVSMESDVATIGSVRADSEALPRMDARTAEINYSKMOKIKSLAFGPDRIEML 703
Db      617 --REDAPQLASDDMTLTGEVAPDAGE--PGLDL-----ALDLPDDDSAEVP 658
Qy      704 PEVLDMGLKIWTDRAEFTQGLGVYDTLLKLSVDSVTYSAD-----GTRALVEATLEE 758
Db      659 PAVPECE-----TPPOSTSAPARSLDPSIDEIDLSGLDPAADAAPASGPAALADWLSLPE 714
Qy      759 SACL--SDLVHP 768
Db      715 QMGLGDDDLAQ 725

RESULT 13
S70368
DNA polymerase I - Bacillus stearothermophilus
C/Species: Bacillus stearothermophilus
C/Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C/Accession: S70368
R/Rigden, M.G.; Tudor, S.; Sivaram, M.; McDonough, S.H.
Biochim. Biophys. Acta 1307, 178-186, 1996
A/Title: Construction of single amino acid substitution mutants of cloned Bacillus stear
A/Reference number: S70368; MUID:96283813; PMID:8679703
A/Accession: S70368
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-876 <RIG>
A/Cross-references: UNIPROT:P52026; EMBL:L42111; NID:9806280; PIDN:AAG37139.1; PID:98062
C/Genetics:
A/Start codon: TTG
A/Superfamily: DNA-directed DNA polymerase I

Query Match      3.2%; Score 128.5; DB 2; Length 876;
Best Local Similarity 21.0%; Pred. No. 3.9;
Matches 170; Conservative 135; Mismatches 302; Indels 201; Gaps 43;
Qy      99 HFLTIDGIRAFEARVSKPQFGSDALISRKQIQAACETLSNPSRRREY-NEGILDE 157
Db      107 HYADADIIIGMAARARE--GFAVKVISGRDLTQ-----LASPVATVEITKKGIRDIE 158
Qy      158 EATVITDV-----PMDKVRG-----ALCVQEGGSEITVILRVG 190
Db      159 SYPTETVVEKYGLTPQIVDLKGLMGKSDNIPGVPGIGERTIVKLLKQFGVENVLASI 218

```

```

Qy      191 EALLKERLPKSFKO--DVLVWMLAFIDVGRDA---MALDPDFTGYFEVEEALKLQE 245
Db      219 DEIKGEKLENLQYRLAL--LSKQLAICRDAPVELTLD--DIVYKGEKREKVALFOE 275
Qy      246 EGASSLAPDLRAQIDE-----TLEETPRVYLLLEIG--PLGDDYAAKRLNG- 290
Db      276 LGFQSPDKMAVQTDDEGEKPLAGMDFAIDSVYDEMADKAAAVVEVVGNDYHNAPIVGI 335
Qy      291 -----LSGVRNLTMSVGGGGAALVQGLTRKFMNDAFLRMTAEE----- 330
Db      336 ALANERGFRLPBTALADKFLAW-----LDDEYKKTMDQSKAAVALMKKGT 385
Qy      331 ----QVDFVATPSNIPAS-----FEVEVALVAQAFIGK-----KPHLQ 370
Db      386 ELRGVFPDLLAAVLDPAAGDVAAVAMGQYEAVRSEAVYGAKARTVDEPTLAE 445
Qy      371 DADKQFOQLQAQVMAWEIPIAMLYDTRNWE-IDFGLERGLCALLIG-----KVDECRM 423
Db      446 HLAIR-----KAAAIWALEEPLMDELRRNEQDRLLTLELPAGILANMEFTGVKVDTKRL 500
Qy      424 -WIGLD--SEDSQYRNPAIYEFVLNSNRDNDLDPGLCKLLETWLAGVVPFRDTRDKK 481
Db      501 EQWGAELTEQLQ---AVERRIYELAQEENINSP---KQLGT---VLFDKLQLPVLKX 549
Qy      482 FKLGDYDDPMVLSYL--ERVEVQGSPLAAATMARIGAEHVKASAMQALQKVPFRYTD 540
Db      550 TKTG-YTSDADVLEKLAHPHIEYE--HILHYRQLGKQSYIE--GLLKVNHPYTGKVNHTM 605
Qy      541 RNSAEPKDVQET--VPSVDVGNNGVGDGEPGVFIABAVRSE-----NETNDYAI 590
Db      606 FNOA-----LTQGRLSVSEPNLQNIPIRLEEGRKIRQAEPSPBDWLIFADYSQJELRV 661
Qy      591 RAGVSESSVDETFE-----MSVANMLKE---ASVKIIAAGVAILISLFSQ 634
Db      662 LAHIAE--DDNILEAFRRGLDITHTKAMDIFHVSEEDVTANRRQAKANFGIIVGISD 718
Qy      635 KYFLKSSSFORKDWVSMESDVATIGSVRADSEALPRMDARTAEINYSKMOKIKSLAF 694
Db      719 -YGLAQNLNITRKEAAEFIRYFASFPQVK-----QYMNIVGE-AKQGYAT 764
Qy      695 GPDRIEMLPEVLDMGLKIWTDRAA-ETQAQGLVYDTLLKLSVD-SVTVSADGTRA- 750
Db      765 TLHRRRYLPDITSRPNFNVSPFAERTAMTPIGSAAAD-IKKAMIDLSVRLREERLQAR 823
Qy      751 -----LVEATLESACLSDLVHPE 769
Db      824 LLLQVHDELLEAPKEIERLCRLV-PE 850

RESULT 14
E83594
Still frameshift probable component of chemotactic signal transduction system PA0413 [lm]
C/Species: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C/Accession: E83594
R/Stover, C.K.; Pham, X.Q.; Edwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Llm,
., Lory, S.; Olsen, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: E83594
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2472 <STO>
A/Cross-references: UNIPROT:Q91696; GB:AE004478; GB:AE004091; NID:9946261; PIDN:AAG0380;
A/Experimental source: strain PA01
C/Genetics:
A/Gene: PA0413

Query Match      3.1%; Score 127.5; DB 2; Length 2472;
Best Local Similarity 20.0%; Pred. No. 23;

```

Matches 159; Conservative 105; Mismatches 291; Indels 239; Gaps 36;

```
124 DALISRRQILQACETLSNPSRREYNGLDDEBEATYITDVPMDKVPGLCVQEGGE-182
19 ETLQARQALFAFENPDPTFRM-----FCITTV--HQVGSTIQWVEFYGA 64
183 --TEIVLRVGEALLKERLPKSFKODVVLVMAIAFLDVSADMAALDPPDF-----ITGYE 234
65 LLAEEMEQVLQALDGRVFNQGEALLEVLMQALILQLPVLDRIQTRARLDPMVVLPLANDL 124
235 FVEBALKLQEGGSSSLAPDL---RAQID-----ETLEETPRVLELGLP 278
125 RAARGEKLLSE--TSLFAPDLISQRPQLDGBAIAQIARTDELGGILRKLRQTOQVALVELL 182
279 LGDDVPAK-----RLNGLSGRNI--LMSVGGGASALVGLTREKFNNEAFLMT 327
183 RNQDVATSLGLIARVYARLEGLCREAPLGLMST---ASGLVEGLANGSVVNSASVA-T 237
328 AAEQVD-----LFVATPSNI-----PAEFVEYEVALLVAQAFTG 363
238 LLARQDLREIKRLVFGADGLNQAAAPDELVKOLLFVAKAPSGSPRIRAL-----286
364 KKPRLQDA--DKGFQQLQAKVMAEIPAM-----LYTRNNWEIDFGLERG--409
287 KEQYRLDEALPDHETVAERARLAGPRDARSVVGALCEELVRIKSLDLFVRSDRGHP 346
410 --LCAAL--IGKVBCRWMLGLDSEDSQYRPAIYEFVLENSNDDND---DLPLGLKL 461
347 SELDALAPLQIADTLAVLGFQGRKYLIDQLDVITHALAQGRBPSAIIIMDVALLY 406
462 LETWLAGVVP--RRDTRDKKKFKLGDDYD-----DPMVLSYL-----497
407 VEATLAGMAGGDERNSEESRLPTTDVAQIHQVITKEARNGLEQAKALIEFIASQNMHE 466
498 -----ERVEVQGS---PLAAATMARIGAHEHVKASAMQALQKVP-----535
467 HLAARPELITQVRGLAMIPLERATTLLETCKRYIQEQL--ARKAVPMQSLDLTADAI 524
536 ---SRVYDRNBAEKDQVEYFSV--DPVGN-----NVRDDEPGVFIAEAVRPSNF 583
525 TSVEYIERLSESDHASQSLDILDVAEDSLANLGLTKENSSAPAPFGI-----572
584 ETNDYAIRAGVSESSVDETTVMESVADMLKEASVKKILAGVAIGLISLFSQYFLKSSS 643
573 -SGRAALIESPAEEREPAAVE--VAETAEQPADTAAEA-----611
644 FQRKQWVSMESDVATIGSVRADSEALPRMDARTAEIVTSKQKIKSLAFGPHRIEM-702
612 --REDAPQLASDDWMTLGEVAPDAGE--PSIDL-----ALDPLDLSAEP 653
703 --LPEVLDGRMLKITDRAAFTAQGLVYDTLKLKLSVSVTSAD-----GTALVEAT 755
654 PALPEVVEBSGQPOSTAPARSLD---DPSLDEIDLSGLDLPADAPAGPALADWS 708
756 LEESACL--SDLVHP 768
709 LPBQWGLGDDLAQP 722
```

RESULT 15

T30853

antigenic heat-stable protein - Rickettsia japonica

C/Species: Rickettsia japonica

C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004

C/Accession: T30853

R/Uchiyama, T.

submitted to the EMBL Data Library, May 1997

A/Description: Sequencing of the gene encoding the antigenic heat-stable 120-kilodalton

A/Reference number: Z20905

A/Accession: T30853

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1018 <UCH>

A/Cross-references: UNIPROT:Q9AJ79; EMBL:AB003696; NID:d1106637; PID:d1020958; PIDN:BAAZI

Query Match 3.1%; Score 126.5; DB 2; Length 1018;

Best local similarity 19.2%; Pred. No. 6.6;

Matches 128; Conservative 109; Mismatches 265; Indels 163; Gaps 28;

```
115 KPPQFGSDALISRRQILQACETLSNPSRREYNGLDDEBEATYITDVPMDKVPGL 174
374 EPPYKQOVPIPTTNOQLP--ETSGWPGS--QGVNPLNLN--AATALSGMQDLNLVYN 428
175 CVLQEGGTEIVLRVGEALLKERLPKSFKODVVLVMA--LAFDVSADMAALDPPDFITG 232
429 AGLRK--KIDLIKEAATAIINDK-----KSDIAEKQANITAIANTVNNKULTPDAKAG 481
233 YEFVEBALKLQEGGSSSLAPDLR-----AQIDETLEETPRVLEL---LGLP 278
482 VNAVLEITKIDN-----TPDLKSKMLEATVVAIALNSEULEKQKQIILEKAVDGLS 535
279 LGDDVPAKRLNGLSGVNIIIMS VGGGASALVGLTREKFNNEAFLR--MTAAEQVDLFV 336
536 LKQD-----ASRAAIDGIT-----DAVKSNIJSTEDKGTWFI 568
337 ATPSNIPAESFEYEVALLVAQAFIGKPHILODADKQFOQLQAKMA-----ME 388
569 AVGDKNVVSLSNAEKQKLGSVLKKGVBAQVLSPAQQLMQNIDKITAQTKKDTIKK 628
389 IPAMLYTRNNWE-----IDFGLERGICALIGKVBCRWMLGLDSEDSQYR 435
629 VNDILFPLSTELKTTNIGAIISNVLDGPATAEVKGIIETITNTYAGSSLEAHD---684
436 NPAIVEFLN--SNRDDNDLPGLCKLLETWLAGVFPFRDTRDKKKFKLGDDYDDPMV 493
685 KAAIIKIGISTIAHSDTSLSPKALIMASAEKGIA-----ESQANLPD--RELMT 734
494 LSYIERVEVVOGSLAAATMARIGAHEHVKASAMQALQK--VFPKRYTDRNSA-----544
735 KGLVDGIYEGKGPEITKAVSSGIDNSININDESEKALKKADAASEALDRETQNLTEGL 794
545 -----EPKD-----VOETVPSVDVPGNNVGRDGEPGVFIAR-----575
795 KQNIIEHKPHDDIYNKRVREYINAVNPIEALSKSPVSAEERIVQETSSIIINISKL 854
576 AVRPSENFETNDYAIRAGVSESSVDETTVMESVADMLKEASVKKILAGVAIGLISLFSQK 635
855 AVEKVNMF-----RAMLSPNGMLKTLIEK-----KESEIKKYDE-----LYKAGTK 896
636 YFLKSSSPQRKQWVS---SMESDVATIGSVRAD--SEALPRMDARTAE--NIVSKQK 688
897 SSTEEQGSFIKANLIDDKTLKSKETRLQITDKLQEQRAEAIENPSVTEDEVVVGSKSK 956
689 IKSIA 693
957 LKPLIS 961
```

Search completed: June 10, 2005, 01:42:11

Job time : 34 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 9, 2005, 21:47:04 ; Search time 95 Seconds

(without alignments)
4317.636 Million cell updates/sec

Title: US-10-600-070-2

Perfect score: 4063
Sequence: 1 MEALSHVIGIGLSPFQLCRLP.....YEVFWSKSGWKITEGSLAS 801

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4063	100.0	801	Q7XAS0	Q7XAS0 arabidopsis
2	4052	99.7	801	Q7XAR9	Q7XAR9 arabidopsis
3	4051	99.7	801	Q9FIC9	Q9FIC9 arabidopsis
4	1775.5	43.7	760	Q6Z729	Q6Z729 oryza sativ
5	1775.5	43.7	760	Q7PC78	Q7PC78 oryza sativ
6	1646	40.5	324	Q7XAS1	Q7XAS1 arabidopsis
7	497	12.2	798	Q8YTL0	Q8YTL0 arabidopsis
8	448.5	11.0	626	Q7NFP3	Q7NFP3 glycochacter
9	442	10.9	631	Q93AK0	Q93AK0 synechococc
10	431	10.6	819	Q8VY16	Q8VY16 arabidopsis
11	429	10.6	714	Q55559	Q55559 synechocyst
12	425	10.5	841	Q91J12	Q91J12 arabidopsis
13	423	10.4	673	Q8DKU7	Q8DKU7 synechococc
14	344.5	8.5	702	Q7V0H1	Q7V0H1 prochloroc
15	340	8.4	653	Q7US57	Q7US57 synechococc
16	318	7.8	685	Q7VNU3	Q7VNU3 prochloroc
17	317	7.8	661	Q7V8U7	Q7V8U7 prochloroc
18	155.5	3.8	5362	Q93155	Q93155 bacillus su
19	148.5	3.7	5363	Q70J29	Q70J29 bacillus su
20	148	3.6	5363	Q6YK40	Q6YK40 bacillus su
21	145	3.6	2297	Q8YLO8	Q8YLO8 arabidopsis
22	143	3.5	2089	Q39478	Q39478 cycloclotella
23	141	3.5	2076	Q6L2Y0	Q6L2Y0 methanococc
24	140.5	3.5	1835	Q6BMT4	Q6BMT4 debaryomyce
25	139	3.4	1024	Q7RYV6	Q7RYV6 neurospora
26	136.5	3.4	1954	Q7MTI8	Q7MTI8 vibrio vuln
27	134	3.3	678	Q8PHZ9	Q8PHZ9 xanthomonas
28	134	3.3	1238	Q9LEP9	Q9LEP9 brassica na
29	133	3.3	1241	Q9UKK5	Q9UKK5 homo sapien
30	132.5	3.3	1011	Q93GJ3	Q93GJ3 rickettsia
31	132.5	3.3	1951	Q8DB34	Q8DB34 vibrio vuln

32	132.5	3.3	2111	1	MCAS MYCBO	Q02251 mycobacteri
33	131.5	3.2	1194	1	DPOL_VZVD	P09252 varicella-z
34	131.5	3.2	1194	1	Q6QCM7	Q6QCM7 human herpe
35	131.5	3.2	2111	2	P96291	P96291 mycobacteri
36	131	3.2	851	2	Q6B8S1	Q6B8S1 debaryomyce
37	130.5	3.2	642	2	Q70CL5	Q70CL5 anopheles g
38	130.5	3.2	876	1	DPOL_BACST	P52026 bacillus su
39	130.5	3.2	1009	2	Q93G19	Q93G19 rickettsia
40	130.5	3.2	1009	2	Q93G12	Q93G12 rickettsia
41	130.5	3.2	1022	1	SCA4_RICCN	Q52658 rickettsia
42	129.5	3.2	1194	2	Q997M9	Q997M9 human herpe
43	129.5	3.2	2381	2	Q7RAR8	Q7RAR8 giardia lam
44	129	3.2	1241	2	Q9BZV3	Q9BZV3 homo sapien
45	129	3.2	1275	2	Q9PEA4	Q9PEA4 neurospora

ALIGNMENTS

RESULT 1

Q7XAS0 PRELIMINARY; PRT; 801 AA.

AC Q7XAS0; 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Division protein.
GN Name=ARC6;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
RX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22779057; PubMed=12897262;
RA Vilha S., Froehlich J.B., Kosharova O., Pyke K.A., Van Ert H.,
Oesteryoung K.W.;
RT "ARC6 is a U-domain plastid division protein and an evolutionary
RT descendant of the cyanobacterial cell division protein Ftn2.";
RL Plant Cell 15:1918-1933(2003).
DR EMBL; AY221468; A018645.1; -
DR InterPro; IPR001623; DnaU_N.
SQ SEQUENCE 801 AA; 88305 MW; D09E27253B46FCC6 CRC64;

Query Match 100.0%; Score 4063; DB 2; Length 801;
Best Local Similarity 100.0%; Pred. No. 7e-241;
Matches 801; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MEALSHVIGIGLSPFQLCRLP	ATTKLRSHNTSTTICSAKMDRLDSEFTSDSSSS	60
Qy	1	MEALSHVIGIGLSPFQLCRLP	ATTKLRSHNTSTTICSAKMDRLDSEFTSDSSSS	60
Db	1	MEALSHVIGIGLSPFQLCRLP	ATTKLRSHNTSTTICSAKMDRLDSEFTSDSSSS	60
Qy	61	FATATTATVSLPPSIDRERHVPIDIFVYVLAQTHFLTDGIRAFARVSKPPQFG	120	
Qy	61	FATATTATVSLPPSIDRERHVPIDIFVYVLAQTHFLTDGIRAFARVSKPPQFG	120	
Db	61	FATATTATVSLPPSIDRERHVPIDIFVYVLAQTHFLTDGIRAFARVSKPPQFG	120	
Qy	121	FSDDALISRQIIOAACETISNPSRREYNEGLDDEAVITDVPMDKYPGALCVIOEG	180	
Qy	121	FSDDALISRQIIOAACETISNPSRREYNEGLDDEAVITDVPMDKYPGALCVIOEG	180	
Db	121	FSDDALISRQIIOAACETISNPSRREYNEGLDDEAVITDVPMDKYPGALCVIOEG	180	
Qy	181	GETREIVRVEGALLKELPKSPKODVYLMMALAFIDVSRDMMALDPDFITGVEFEAL	240	
Qy	181	GETREIVRVEGALLKELPKSPKODVYLMMALAFIDVSRDMMALDPDFITGVEFEAL	240	
Db	181	GETREIVRVEGALLKELPKSPKODVYLMMALAFIDVSRDMMALDPDFITGVEFEAL	240	
Qy	241	KLQEBGASSIAPDLRAQIDETLEITPRYVLELGLPLGDDYAAKRLNGISGVNIIWS	300	
Qy	241	KLQEBGASSIAPDLRAQIDETLEITPRYVLELGLPLGDDYAAKRLNGISGVNIIWS	300	
Db	241	KLQEBGASSIAPDLRAQIDETLEITPRYVLELGLPLGDDYAAKRLNGISGVNIIWS	300	
Qy	301	VGGGASALVGLTRKFNNEAFLRMTRAAOVULFVATPSNIPESFEYVEVALVAQA	360	
Qy	301	VGGGASALVGLTRKFNNEAFLRMTRAAOVULFVATPSNIPESFEYVEVALVAQA	360	
Db	301	VGGGASALVGLTRKFNNEAFLRMTRAAOVULFVATPSNIPESFEYVEVALVAQA	360	

```

QY 361 FIGKKPHLQDADKQFOOLQOAKVMAEIPAMLYDTRNNMEIDFGELGICALLIGKVD 420
DB 361 FIGKKPHLQDADKQFOOLQOAKVMAEIPAMLYDTRNNMEIDFGELGICALLIGKVD 420
QY 421 CRMWLGIDSDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVPPRRDTKDK 480
DB 421 CRMWLGIDSDSQYRNPAIVEFVLENSNRDNDLPGCLKLETWLAGVPPRRDTKDK 480
QY 481 KFKLGDYDDPMTLSYLERVEVQSSPLAAATMARI GAHVKASAMQALQKVPSSRYTD 540
DB 481 KFKLGDYDDPMTLSYLERVEVQSSPLAAATMARI GAHVKASAMQALQKVPSSRYTD 540
QY 541 RNSAEPDQVETVFSVDPVGNNGRDEPGVFIAEAVRPSSENFETNDYAIRAGVSSSSVD 600
DB 541 RNSAEPDQVETVFSVDPVGNNGRDEPGVFIAEAVRPSSENFETNDYAIRAGVSSSSVD 600
QY 601 ETTVEMSVADMKEASVKIIAAGVAIGLISLFSQKFLKSSSFQRKDMVSMESDVATI 660
DB 601 ETTVEMSVADMKEASVKIIAAGVAIGLISLFSQKFLKSSSFQRKDMVSMESDVATI 660
QY 661 GSVRADSEALPRMDARTAEIVSKMOKISLAFGPHRIEMLPVLDGMLKIWDRAA 720
DB 661 GSVRADSEALPRMDARTAEIVSKMOKISLAFGPHRIEMLPVLDGMLKIWDRAA 720
QY 721 ETALQGLVYDYLKLSVDSVTSVADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
DB 721 ETALQGLVYDYLKLSVDSVTSVADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
QY 781 RYEVFMSKSGMKITEGSVLAS 801
DB 781 RYEVFMSKSGMKITEGSVLAS 801

```

RESULT 2

```

QYXAR9 ID QYXAR9 PRELIMINARY; PRT; 801 AA.
AC QYXAR9;
DT 01-OCT-2003 (TREMBLrel. 25, Created)
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Division protein.
GN Name=AKC6;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22779057; PubMed=12897262;
RA Vitha S., Froehlich J.E., Kosharova O., Pyke K.A., Van Eyr H.,
RA Oesteyoung K.W.;
RT "ARCE is a J-domain plastid division protein and an evolutionary
RT descendant of the cyanobacterial cell division protein Ftn2.";
RL Plant Cell 15:1918-1933(2003).
DR EMBL; AY221469; AAC18646.1; -.
DR InterPro; IPR001623; DnaJ N.
SQ SEQUENCE 801 AA; 88247 MW; 7E2E1B3FD4B61 CRC64;

Query Match 99.7%; Score 4052; DB 2; Length 801;
Best Local Similarity 99.8%; Pred. No. 3.3e-240;
Matches 799; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MEALSHVIGLSPPQCLRLPAPATTKLRSHNTSTTISASAMDRLLSDNFTSDSSSS 60
DB 1 MEALSHVIGLSPPQCLRLPAPATTKLRSHNTSTTISASAMDRLLSDNFTSDSSSS 60
QY 61 FATATTATVSLPPSIDRERHVPIDPFYQVLAQTHFLTDGIRAFKARVSKPPQFG 120
DB 61 FATATTATVSLPPSIDRERHVPIDPFYQVLAQTHFLTDGIRAFKARVSKPPQFG 120
QY 121 FSDALISRRQIIQAACETLSNPRSREYNGLIDDEATVITDVPMDKVPALCVLQEG 180
DB 121 FSDALISRRQIIQAACETLSNPRSREYNGLIDDEATVITDVPMDKVPALCVLQEG 180

```

RESULT 3

```

QYFTG9 ID QYFTG9 PRELIMINARY; PRT; 801 AA.
AC QYFTG9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 5, Pl clone:MDH9
DE (Hypothetical protein At5g42480).
GN Name=At5g42480;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99156233; PubMed=10048486;
RA Aamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT Sequence features of the regions of 1,081,958 bp covered by seventeen
RT physically assigned Pl and TAC clones.";
RL DNA Res. 5:379-391(1998).
RN [2]
RP SEQUENCE FROM N.A.

```

Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
 Dang J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
 RA Dang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
 RA Davis R.W., Ecker J.R., Theologis A.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Huanan V.W., Lee J.M.,
 RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
 RA Southwick A., Tzipf M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB016888; BAB10489.1; -
 DR EMBL: AY091075; AAM13895.1; -
 DR EMBL: AY150490; AAN12907.1; -
 DR InterPro: IPR001623; DnaJ_N.
 KW Hypothetical protein.
 SQ SEQUENCE 801 AA; 88259 MW; 608E776FBA73FECF CRC64;
 Query Match 99.7%; Score 4051; DB 2; Length 801;
 Best Local Similarity 99.8%; Pred. No. 3.8e-240;
 Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MEALSHVIGLSPQOLCRLPATTKLRSHNTSTTIC-SASKMADRLSDPFTSDSSSS 60
 DB 1 MEALSHVIGLSPQOLCRLPATTKLRSHNTSTTIC-SASKMADRLSDPFTSDSSSS 60
 QY 61 FATATTTATVLSLPPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFG 120
 DB 61 FATATTTATVLSLPPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFG 120
 QY 121 PSSDALISRQIIOACETLSNPRSRRYNEGGLDDEATYITVPMDKVCALCVLOEG 180
 DB 121 PSSDALISRQIIOACETLSNPRSRRYNEGGLDDEATYITVPMDKVCALCVLOEG 180
 QY 181 GETHIVLRVGEALKEKLPKSKFQDVVLMALAFIDVSRDAMALDPDFITGYEFVEBAL 240
 DB 181 GETHIVLRVGEALKEKLPKSKFQDVVLMALAFIDVSRDAMALDPDFITGYEFVEBAL 240
 QY 241 KILQEBGASLAPDLRAQIDETLEITPRVYVLELGLPLGDDVAAKRLNGLSGVNITLMS 300
 DB 241 KILQEBGASLAPDLRAQIDETLEITPRVYVLELGLPLGDDVAAKRLNGLSGVNITLMS 300
 QY 301 VGGGGAALVGGLTREKFMNEAFPLMTAAEQVDLPVATPSNIPASFEVYVALVAQA 360
 DB 301 VGGGGAALVGGLTREKFMNEAFPLMTAAEQVDLPVATPSNIPASFEVYVALVAQA 360
 QY 361 FIGGKPHILOADADKOFQOLQOAKVMAEIPAMLYTRNNWEIDFGLEBGLCALLIGKVD 420
 DB 361 FIGGKPHILOADADKOFQOLQOAKVMAEIPAMLYTRNNWEIDFGLEBGLCALLIGKVD 420
 QY 421 CRMTGLDSEDSQYRNPAIVEFVLENSNRDNDLPGICKLETLWLAGVPPRFDTDX 480
 DB 421 CRMTGLDSEDSQYRNPAIVEFVLENSNRDNDLPGICKLETLWLAGVPPRFDTDX 480
 QY 481 KFKLGDYDDPMVLSYERVAVVGGSPAAAAAARICAEHYKASAMQALQKVPSTTD 540
 DB 481 KFKLGDYDDPMVLSYERVAVVGGSPAAAAAARICAEHYKASAMQALQKVPSTTD 540
 QY 541 RNSAPKPOVOETVSVPGNNVGRDGEVFIABAVPSENFTNDPAIRAGSESSVD 600
 DB 541 RNSAPKPOVOETVSVPGNNVGRDGEVFIABAVPSENFTNDPAIRAGSESSVD 600
 QY 601 ETTVEMSVADMLKEASVYKILAAVAGLISLFSQKTFPKSSSSFORKDMVSMSDVATI 660
 DB 601 ETTVEMSVADMLKEASVYKILAAVAGLISLFSQKTFPKSSSSFORKDMVSMSDVATI 660

GSVADSEALPRMDARTAEIVSKMOKIKSLAGPDRHLEMLPEVLDEGRLKITWDRAA 720
 DB 661 GSVADSEALPRMDARTAEIVSKMOKIKSLAGPDRHLEMLPEVLDEGRLKITWDRAA 720
 QY 721 ETTAQLGVYDYTLTKLSVDSVTSVADSTRALVEATLEESACLSDLVHPENNATDVRTYT 780
 DB 721 ETTAQLGVYDYTLTKLSVDSVTSVADSTRALVEATLEESACLSDLVHPENNATDVRTYT 780
 QY 781 RYEVFWSKGMKITEGSLAS 801
 DB 781 RYEVFWSKGMKITEGSLAS 801
 RESULT 4
 ID 062729 PRELIMINARY; PRT; 760 AA.
 AC 062729;
 DT 05-JUL-2004 (TREMBLrel. 27, Created)
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
 DE Plastid division protein.
 GN Name=P0575F10.2;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Euphorbiaceae; Oryzae; Oryza.
 CX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa japonbare (GA3) genomic DNA, chromosome 2, PAC
 clone: P0575F10.2";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP004885; BAD07942.1; -
 DR InterPro: IPR001623; DnaJ_N.
 SQ SEQUENCE 760 AA; 84134 MW; 2C44684862795B2F CRC64;
 Query Match 43.7%; Score 1775.5; DB 2; Length 760;
 Best Local Similarity 47.9%; Pred. No. 2.1e-100;
 Matches 390; Conservative 119; Mismatches 213; Indels 93; Gaps 16;
 QY 12 SPFQCLRPATTKLRSHNTSTTIC-SASKMADRLSDPFTSDSSSSPATAT 65
 DB 14 APFAPSLPRPRPRRRRPPPHPSAACRAASRMARLTADPHLPTAAPSDPPSAPAPAA 73
 QY 66 TTATVLSLPPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFGSDA 125
 DB 74 APSAPSPVPLPPDAABERSLPQVDVFKVLAEPHPLDGGIRRAFEARVAKPPQYGVSTDA 133
 QY 126 LISRQILOACETLSNPRSRRYNEGGLDDEATYITVPMDKVCALCVLOEGGETEI 185
 DB 134 LVGRROMQLAHDITLQNRSRTQYDRALSENREBALTMIDAMDR-----SAGEALA 184
 QY 186 VLRGEBALKEKLPKSKFQDVVLMALAFIDVSRDAMALDPDFITGYEFVEBALKILQ 245
 DB 185 VLTVGEQLLDRPKRKQDVVLMALAYVDSRDAMAAAPVDVIGCCFVLERLKLQ 244
 QY 246 EGASLAPDLRAQIDETLEITPRVYVLELGLPLGDDVAAKRLNGLSGVNITLMSVGGG 305
 DB 245 DGASNLAPDLISQIDETLEITPRCVLELGLPLDTHKKRKGGLGAGRYITLMSVGRG 304
 QY 306 ASALVGLTREKFMNEAFPLMTAAEQVDLPVATPSNIPASFEVYVALVAQAFIGKK 365
 DB 305 IATVGGFSEAFNMERFLMTSIEQMDFPSKTPNSIPPEWFEIYVALVAQAIIISK 364
 QY 366 PHLQADKOFQOLQOAKVMAEIPAMLYTRNNWEIDFGLEBGLCALLIGKVDCEKML 425
 DB 365 PPIIMADDLPEQOKNIGS----HYAYDN----ENDLALERFCSGLIVDVSKCMM 416
 QY 426 GLDSEDSQYRNPAIVEFVLENSN-RDNDNDLPGICKLETLWLAGVPPRFDTDXKFKL 484
 DB 417 GIDVSESPYDPLKEIFLTVNNSISEENDLPGICKLETLWLEFEVPPRSRDTGMOFRL 476

[illegible]

ID	Q7PC78	PRELIMINARY;	PRT;	760 AA.
AC	Q7PC78;			
DT	01-MAR-2004 (TREMBLrel. 26, Created)			
DT	01-MAR-2004 (TREMBLrel. 26, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	Plastid division protein.			
GN	Name=Arc6;			
OS	Oryza sativa (indica cultivar-group).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyte; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Ethnartoidae; Oryzae; Oryza.			
OX	NCBI_TaxID=39946;			
RN	[1]			
RP	SEQUENCE FROM N. A.			
RX	MEDLINE=22779057; PubMed=12897262;			
RA	Velha S., Froehlich J.E., Koshnareva O., Pyke K.A., Van Erp H.,			
RA	Osteeroung K.W.;			
RT	"ARC6 is a D-domain plastid division protein and an evolutionary			
RL	descendant of the cyanobacterial cell division protein Fz2.";			
RL	Plant Cell 15:1918-1933(2003).			
CC	-I- MISCELLANEOUS: The sequence shown here is derived from an			
CC	EMBL/genbank/DDBJ third party annotation (TPA) entry.			
DR	EMBL; BK000999; DAA01472.1; ..			
DR	Gramene; Q7PC78; ..			
DR	InterPro; IPR001623; DnaJ_N.			
SQ	SEQUENCE 760 AA; 84134 MW; 2C4684862795B2F CRC64;			
Query Match 43.7%; Score 1775.5; DB 2; Length 760;				
Best Local Similarity 47.9%; Pred. No.21e-100;				
Matches 390; Conservative 119; Mismatches 213; Indels 93; Gaps 16				
QY	12 SPFLQRLPPATTKLRSHNTSTTC-SASKWADRLSDNF-----TSDSSSSSFAVAT 65			
Db	14 APFAFSLPRPRPRRRRPPHPSAACRAASRWAERLFADPHLLPTPAASDPSPAPADAA 73			
QY	66 TTALTVLSPEIDREHVRPIPIIDYOVYGAQTHRTGIGRAFAVARVSKPGFSDA 125			
Db	74 APSASFPVPLPDDAERSLPQVDDYKYKLGABPHVLSGIRAFARAKAPPOYVSTDA 133			
QY	126 LISRQLLOAACETLSNPSRSREYNEGLLDDEEATVTDVPMDKVPALCVLOEGETEI 185			
Db	134 LVGRKWLQIHADTLTMNQNSRTOYDRALSENREBALTMIDAMDK-----EAGGALA 184			
QY	186 VLRGEALIKERLPSKFDYDVLVWALAFLDVSRDAMALDDPPDFTGYEYFEALIKLQ 245			

Db	189	VLTGSGQLLDBRPPKRFQDQVVLAMALAYVDLSRDNAMASPPDVIYGCCEVLERALXULQE	244
Qy	246	EGASSIAPRLQAOIDTELTETTPRYVYLIGLPLGDDYAAKRLNGSGVRNIIWSVGGG	305
Db	245	DGASNIAPRLLSQIDETTEETETPRCVLELISPIIDTEHHKROGLOGARINILMSVORG	304
Qy	306	ASALVGLTRKPKMNEAPLRMTAAEOVDLPVATSGNTPAESFEFEYEVALLVQAQFGKK	365
Db	305	IATVGGGFSEBAPMNEAPLRMTSIEOMPFSKTPISLPDPEFELIYNALHVAOALISKR	364
Qy	366	PHLLQADAPQFOOLOQAKYAMEIRPAMLYDTRNNWEIDFGELGICALLIGYDECRMWL	425
Db	365	PQITMAADLPLQLOKFNIGS----HYAYDN----BMDLALERAFCSILVDGYSKCMWL	416
Qy	426	GLDSESSQTRNPALVEFVLNSN-RDDNDLPLGLCKLLETMLAGVPPRRDRTOKKKFKL	484
Db	417	GIDNESSPRDRKILFEIYVNTSSISEENDLPLGCKLETMLIPEVPRSRDRGMQFRL	476
Qy	485	GDYDDPMTLSLERVEVQSGPLAAATMMIRIGAE-----HYKASAMQALOKFVP-SR	537
Db	477	GDYDDPEVLTSLERVEGGASHLAAAAIKLQAOATALGTVYKSNIAQAFNKFVPLIE	536
Qy	538	YTDNRSAEPKDOETVFSVDPVGNVNGDGEPPGVFIABVAPSENPEFTNDYAIRAGVSES	597
Db	537	QLDRSAMENT-----KDG-PGGYL-----ENFDQ-----	559
Qy	598	SUDETTEVMSVADMLKEASVKILAGVALIGLISLPSQKFLKSSSSSQORDMSSMSMD-	656
Db	560	-----ENAPPHDSRNALKIIISGALFALLAIGALY-----LPRRPPISALARSEH	605
Qy	657	--VATIGSVRADDSAL-----PRMDARTAEINIVSMQKIKSLAEGPHRIEMLEPEVL	707
Db	606	GSAYAVANSIDSDDDPALDEDPVHITPRMNAKLAEDIYRKWQSIKSKALGPEHSVANSLOEVL	665
Qy	708	DGRMLKITWDRAAETAOGLVYDYTLKLSDVSATVSSADGTRALVEATLEESACLSOLVH	767
Db	666	DGNMLKVTWDRAAABIERHGMFEWETYLSDVTIDISLIDSGRRATVEATIDEAQGLFDVTE	725
Qy	768	PENNATDVATTTTRYEVFMASK-SGKIKTEGSVLAS	801
Db	726	PRNDSYDTRKYTRYEMAFSKLGGKMIKEGAVLAS	760

RESULT 6			
07XAS1			
ID	07XAS1	PRELIMINARY;	PRT; 324 AA.
AC	07XAS1;		
DT	01-OCT-2003 (TREMBLrel. 25, Created)		
DT	01-OCT-2003 (TREMBLrel. 25, Last sequence update)		
DT	01-MAR-2004 (TREMBLrel. 25, Last annotation update)		
DE	Truncated division protein.		
GN	Name=arc6;		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22779057; PubMed=12897262;		
RA	Viltha S., Froehlich J.E., Koksharova O., Pyke K.A., Van Erp H.,		
RA	Osteryoung K.W.;		
RT	"ARC6 is a J-domain plastid division protein and an evolutionary		
RT	descendant of the cyanobacterial cell division protein Ftn2.";		
RL	Plant Cell 15:1918-1933 (2003).		
DR	EMBL; AY221467; AAQ18644.1; -.		
DR	InterPro; IPR001623; DnaJ_N.		
SQ	SEQUENCE	324 AA;	35468 MW; OCC5C0CC469A325 CRC64;
Query Match 40.5%; Score 1646; DB 2; Length 324;			
Best Local Similarity 100.0%; Freq. No. 5,4e-93;			
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			

```

QY 1 MEALSHVIGISLPOLGRLPATTKLRSHNTSTTTICGASKMADRLSDENFTSDSSSS 60
Db 1 MEALSHVIGISLPOLGRLPATTKLRSHNTSTTTICGASKMADRLSDENFTSDSSSS 60
QY 61 PATATTTATVSLPSPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFG 120
Db 61 PATATTTATVSLPSPSIDRPERHVPIDFYQVLAQTHFLTDGIRRAFEARVSKPPQFG 120
QY 121 FSDALISRRQIIQAACETTSPRSRRYNEGILLDDEAATITVPMVKVGCALCVLOEG 180
Db 121 FSDALISRRQIIQAACETTSPRSRRYNEGILLDDEAATITVPMVKVGCALCVLOEG 180
QY 181 GETEIVLRVGEALIKERLPKSPKODVVLVMAALFLDVSADAMALDPDFITGEFEVEAL 240
Db 181 GETEIVLRVGEALIKERLPKSPKODVVLVMAALFLDVSADAMALDPDFITGEFEVEAL 240
QY 241 KLIQEBGASSIAPDLRAQIDETLEITPRVYVLELGLPLGDDVAAKRLNGISVRNIIWS 300
Db 241 KLIQEBGASSIAPDLRAQIDETLEITPRVYVLELGLPLGDDVAAKRLNGISVRNIIWS 300
QY 301 VGGGASALVGLTRKEMNEAFL 324
Db 301 VGGGASALVGLTRKEMNEAFL 324

```

RESULT 7

```

QBYTLO PRELIMINARY; PRT; 798 AA.
AC QBYTLO;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE A112707 protein.
GN OrderedLocustNames=a112707;
OS Anabaena sp. (Strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxId=103690;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Matanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.,
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
RX EMBL; AP003590; BAB74406.1; -.
DR PIR; AD2144; AD2144.
DR HSP; P08622; IBOZ.
DR Pfam; PF00226; DnaJ. 1.
DR SMART; SM00271; DnaJ. 1.
DR PROSITE; PS50076; DnaJ_2; 1.
KM Complete proteome.
SQ SSOURCE 798 AA; 90124 MM; 3C57B66C21722AA CRC64;

```

Query Match 12.2%; Score 497; DB 2; Length 798;

Best Local Similarity 24.6%; Pred. No. 7.9e-22;

Matches 207; Conservative 139; Mismatches 301; Indels 196; Gaps 33;

```

QY 84 VPIPIDPVYVLAQTHFLTDGIRRAFEARVSKPPQFGSDALISRRQIIQAACETTSPN 143
Db 11 VPIPIDPVYVLAQTHFLTDGIRRAFEARVSKPPQFGSDALISRRQIIQAACETTSPN 143
QY 144 RSRREY-----NEGILLDDEAATITVPMVKVGCALCV 176
Db 71 KERSYQVIVLAHAYPDNATTKVAVENRGDSNNGHFDVQSLSI--EVSSEELIGALLI 128
QY 177 LOEGGEIVLRVGEAL-----KERLPKSP-----KQDVVLVMAALFLDVS 219
Db 129 LOEGGEIVLRVGEAL-----KERLPKSP-----KQDVVLVMAALFLDVS 219

```

```

QY 220 D-----AMALDPDFITGEFEVEALKLIQEBGASSIAPDLRAQIDETLEITPR 269
Db 189 EQMGQGHYENALSL-----TQGE-----VLFSSG-----IFPSVQABIQADLYKLRPY 233
QY 270 YLELLPLGDDVAAKRLNGISVRNIIWSVGG--GASALVGLTVE--KFMNEAFLR 325
Db 234 RILELLALP--OEKTIERRHGLDLOSILDRGIDTGTNDOSGLNIDDFLRFIOQLRH 291
QY 326 MTAEOVDLVANPSPNIPASFEVEVALVQAIFIGKKPHILODADKQFOQLQOAKM 385
Db 292 LTVABQKLPDG--SKRPS--AVATYLAIVASIAQFQROPALIRHAKOILMRLSQO-- 347
QY 386 AMEIPAMLYDTRNNWEIDFGLERGLCALLIGKVDCEHMLGLDSEDSQYRNPAIVEVLE 445
Db 348 -----DVHLEGLCALLIGQTEARVLELSE--YEALALI-----R 383
QY 446 NSNRDNDLLPGLCKLETLVAGVPPFRPDYDKKFKLGDYDDPNVLSYLEV----- 500
Db 384 EKSQSDPDLPLGLCLVAEQWLONEVFPHFRDLISROQASLKDYFANQOVQALTEALPDAE 443
QY 501 -----EVQGSPLAAATMAR--IGAEH--VKASAMQALQKVPSP--RYTD 540
Db 444 TTNEWAVINQSPSQPRGNSYSGGTPVAKRPVGKANRGEASTRPVQKSHSPSVNQPH 503
QY 541 RNSAPKQVOETVFSVDPVGNVGRDGE-----PGVFIAAVPSENFEITND 587
Db 504 QNRTPDELPETSHRHPBESSNFTTARENISTTDAYTDNYPPEIPVERASP----- 555
QY 588 YAIRAGVS--ESSVDFTV-----EWSVADMLKEASVKILAA 622
Db 556 --VQPGVSGTQSTPRPQTPKRRRKKPQAVNRGSHIHQOROPSPSTLRKTLRILVL 613
QY 623 GVAIGLSLF-----SQKX-FLKS-----SSSFQRDMVSNESVDATGSRADSEALPR 673
Db 614 G-SLGGILFLVLVSTTFGWLKVNFFPAPSLQSGQLSIQSQPLEIPDKNAQIQSPBVS 672
QY 674 MDARTAEINYSKQKISLAFGPDRIEMLEPEVLGRLKIMTDRRAETAQGLVYDYL 733
Db 673 LTERTAKITENMLATGASLGAHEKIESLEILITGSALQOWRLIALQDADRNRHREYS- 731
QY 734 LKLSVDSVTVS-ADGTRALVEATLESACISDLVH--PENNAVDVTRYTRYEVFWSKSG 790
Db 732 HSYVDSISKSIDIDPNASVGATVRE--LTQFYENGQKKSBER-LRVRYELIRDDDI 787
QY 791 WKI 793
Db 788 WRI 790

```

RESULT 8

```

QBYTLO PRELIMINARY; PRT; 626 AA.
AC QBYTLO;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE G13481 protein.
GN OrderedLocustNames=g13481;
OS Gloeobacter violaceus.
OC Bacteria; Cyanobacteria; Chroococcales; Gloeobacter.
OX NCBI_TaxId=33072;
RN [1]
RP SEQUENCE FROM N. A.
RX STRAIN=PCC 7421.
RX MEDLINE=22977040; PubMed=14621292;
RA Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyashita H., Tsuchiya T.,
RA Sasamoto S., Matanabe A., Kawashima K., Kishida Y., Kiyokawa C.,
RA Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpō S.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of Gloeobacter violaceus PCC 7421, a
RT cyanobacterium that lacks thylakoids."
RL DNA Res. 10:137-145(2003).
RX EMBL; AP006580; BAC91422.1; -.

```


Db 582 AVQAKVDE----VEQYRGDQLLETTRDGLVIRYQLVRENNIMKIASISLV 630

RESULT 10

Q8VY16 PRELIMINARY; PRT; 819 AA.

AC Q8VY16; 01-MAR-2002 (Tremblrel. 20, Created)

DT 01-MAR-2002 (Tremblrel. 20, last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, last annotation update)

DE Hypothetical protein Ac3g19180.

GN Name=Ac3g19180;

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.

OC NCBI_TaxID=3702;

RN [1]

RA SEQUENCE FROM N.A.

RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,

RA Goldenrich A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,

RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,

RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,

RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,

RA Seki M., Shimizu P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,

RA Theologis A.;

RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY074283; AAL6980.1; -

KW Hypothetical protein.

SO SEQUENCE 819 AA; 90750 MW; 6720083F4864657E CRC64;

Query Match 10.6%; Score 431; DB 2; Length 819;

Best Local Similarity 23.5%; Pred. No. 9.4e-18;

Matches 200; Conservative 116; Mismatches 311; Indels 224; Gaps 32;

QY 56 SSSSFATATTATVLSLPPSIDRPERHVPIDIFYOVLAGNQTHTDGRFAFEAVSK 115

DB 81 SRTSSLAASST-----ILPVCYCLIGSEGADEVKSYINLKT 124

QY 116 PPGGFSDDLISRROQLAACETLSNPSRREYEGLLDEEATVTTDVPWKPGALC 175

DB 125 DAEGBYMEAAARQDLMDVDRDL--LPESEYAGNLEKELAKPSLRIPMWTLPALC 181

QY 176 VQEGGTEIVLRGELKERLPKSPKQVYVMAALFDVSDAMALDPDPTGYEF 235

DB 182 LLDVGGEKVLVDIGRALRLDSKPYIHDFLSMALBSCAIAAEEVNVV--QGEFA 239

QY 236 VEEALKIQEE-GASLAPDLRAQIDETLEBITPRVYELTGLPLGDVAKRLNGSGV 294

DB 240 LAAAGSLKSKVTLGKLA--LITQIESLEGLAPCTILDGLPRTENMERKGAIAL 297

QY 295 RNILMSVGGGASALVGLTRK-----FNNEAFLEMTAAEOYDLF-----VATP 339

DB 298 RELLRQ-----GLSVEABSCQIQDWPCFLSQAISRLALTEIYDLLEWDDLATRK 346

QY 340 SNIPABEF-----YEVALLVAQAFIGKPHLDQADQFQLOQAKMAWEIP 390

DB 347 NKSLSESHNORVVIDNFCFTVWLLGHIAVFGS-----KQNETINAKTICECLI 396

QY 391 AMLYDTNNMEIDPGLERGLCALTIKGVDECRMMVGL-----DSEDSQYRNPAIYEVLEN 446

DB 397 A-----SEGVDLKFEAPCSFLLKQSSBALEKQLQESNDSAVRNS-----ILGK 444

QY 447 SNRDDNDLPGCLKLTWTLAGVPPRPRTK-----DKPKGLGDYDPMVL 494

DB 445 ESRSTS-----ATPSLEAMLESVLANFPDTRGCSPLANFFRAEKYKPKKMGKSPS 499

QY 495 SY-----LAEVYVQSSPLAAATMARIGAEHTKASAMQALQVPSRYTDRKSAPKO 548

DB 500 NHTNQRPPLSTQFVNS-----QHL-----YAAVEQLTPTD 531

QY 549 VQETVFPVDVGNVNGDGEKGVFLAEVPSSENFETNDYAIRAGVESSV-DETVEMS 607

Db 532 LQSEVVSAAK--NNDE-----TSASMPVQLKEN-----LGVAKNKIMDEM--LS 571

QY 608 VADMLEKASVKKLLAAGVAILGLISFSQKY----- 636

Db 572 QSSLIGRVSVAL-----LQCTVFFSLKLSGINSGLQSPISVSAAPHSESDSFLMKT 626

QY 637 ---FLKSSSFORKDVS-----MESDVATIGSVRADSEALPR 673

Db 627 SGNFRKLLDSVNNNGIYGNIKVLIDMLKMGHGEHPALYLKSSQASATLSHSASELHKR 686

QY 674 -MDARTENIVSKQKQKSLAFGPDRILEMPEYLDGRMLKITDRALETRQLGLV-DY 731

Db 687 PMDTEBAEELVQWENYKAAALGTHQVYLSLSEVLDESMTVQW-QTLAQTAARAKSCWRP 745

QY 732 TLKLKLSVDSVTSAD--GTRALVEATLESASACSLVHPENNATDVRTTYTRVEFMSK 788

Db 746 VLLHLEVLQAHIFEDGIAGANAALTEALBBAELVDSQPK-NKTYSTYKIRIYLKQ 804

QY 789 SG-WKITEGSV 798

Db 805 DGLMKFCQSDI 815

RESULT 11

Q5559 PRELIMINARY; PRT; 714 AA.

AC Q5559; 01-NOV-1996 (Tremblrel. 01, Created)

DT 01-NOV-1996 (Tremblrel. 01, last sequence update)

DT 01-JUN-2003 (Tremblrel. 24, last annotation update)

DE S110169 protein.

GN OrderedLocustNames=s110169;

OS Synecchocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synecchocystis.

OC NCBI_TaxID=1148;

RN [1]

RA SEQUENCE FROM N.A.

RC STRAIN=PCC6803;

RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kocani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hirosewa M., Sugitani M., Sasamoto S., Kimura T.,

RA Hoshouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shimizu S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synecchocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions."

RL DNA Res. 3:109-136(1996).

DR EMBL; D63999; BAA10060.1; -

DR PIR; S76082; S76082.

DR HSP; P08622; I802.

DR InterPro; IPR001623; DnaJ_N.

DR Pfam; PF00226; DnaJ_1.

DR SMART; SM00271; DnaJ_1.

DR PROSITE; PSS0076; DnaJ_2; 1.

KW Complete proteome.

SO SEQUENCE 714 AA; 79423 MW; 51B52C16F405ED3E CRC64;

Query Match 10.6%; Score 429; DB 2; Length 714;

Best Local Similarity 25.1%; Pred. No. 1e-17;

Matches 194; Conservative 124; Mismatches 331; Indels 124; Gaps 30;

QY 86 IPIDFYVGLAGNTHFLTDGIRRAEAVSKPPQGSDDDLISRQQLQACETLSNPS 145

DB 3 IPIDFYRILIPQSGGETIEQAYODRLQLPRREFSDAAVTYLNQLLATAYETLRDPEK 62

QY 146 RREYNE--GLDDE--EATVITDVPMDKVP-----GALCVLQEGGETEIVLRGEALLKE 196

DB 63 RQAYDQEMKAMBALEALPITTPLEKCPBEQIGALLILDGIEVLVVKGEPTLHD 122

QY 197 RLPKS--FKQDVVLMALAFLDVSRDAMALDPDPTGYEFVEEA-LKLLQEGASSLAP 253

DB 123 PNPAGGLPDYLLSVILAHMELSRERWQQQ-----YEFATATSIKALARLQODNDP 176

```

QY 254 DLRAQIDETLEETPRVYLELGLP-LGDDYAAKRLGSGVNRILMSVGG-GGASALVG 311
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 177 ALBAERHQELRYLRPRYILLELAKEGGGE---QROGGLALLQMWODRGIEGKSDYS 233
QY 312 GLTRE---KEMNEAFLEMTAAEQVDLFVATPNSIPASFEVYALALVAQATIGKPEHL 368
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 GLANDDFLKKFTHORCHLTVAEQNALPL--PESQRPVLASVLAHSLMEGVEKEDPMA 291
QY 369 LODADKQFOOLQOAKVAMEIPAMLVYTRNNWEIDFGLERGLCALLIGKDECMWGLD 428
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 292 IYAKSLIILQLENCQ-----DLLEKVICELLGQTEV--VLAID 330
QY 429 SEDSQYRNPAIVEFVLENSNRDDNDLPGLCKLETWLAGVPEPRFDYOKKFKLGDY 488
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 331 QGD-----PKIVA-GESKLTATGEDPLTAFTYFTEQULBEEIVYFRDLSPETLSPKAYF 384
QY 489 DDMVNTLSYLERVEVQ-----GSP-LAAATMARIGAEHVAKASAM-QALQKVPFRYT 539
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 385 NNPSVOOYLEQLBPSDFTTNSFASPALLTSTATESETPMVHSSAALPDRPLSTVPS--- 441
QY 540 DRMSABPKDQERYFVSVDVPGNNVGRDBPGVFIAEAVRSENETNDVAIRA----- 592
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 442 -RRGRSPRRSRDVF-----SADNSSGLAVT-TLSPALAYDHSIGTGIGGDSTS 491
QY 593 -GVSESSVDETTYE-MSVADMLKEASVK-----ILAAVAILGISLFSQK----- 635
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 492 NGSSNSAPSTGKHSRRRKRVTIKPVRFGLFLCLAGIVGAGATALLINRTGDPDGG 551
QY 636 -----YFLKSSSSFOKCMVSMESDVATIGSVRADDEBALPRMARPAENIVSKWK 688
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 552 LLEDPDLVFL-----DQPSFPIPEKATSRNLISQ---FNFOQVQGVWVGMD 598
QY 689 IKSLAFQPDRIEMLEPVLDGRMLKITDRAETPAQGLVYDYTLKLSTDSVTVS-ADG 747
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 SKGLAFQGNVDVAGLSVLAENLLAQOGR-AQRDQAKYTHQYHNLQTLAYQVNPDR 657
QY 748 TRALVEATLEESACLSDLVHPENNATDVR-TYTTTYEVFMSKSGWKITEGSVL 799
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 658 NRATVTVARVEISQPFILGNQOQSATKODLTVRYQLVBRHQGVWKIDQLQV 710

```

RESULT 12

```

Q9LJL2 PRELIMINARY; PRT; 841 AA.
ID 09LJL2
AC 09LJL2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 15, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MVI11.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI TaxID=3702;
OX NCBI TaxID=3702;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
RN 12
RP SEQUENCE FROM N.A.
RA Kaneo T., Kato S., Nakamura Y., Asamizu E., Tabata S.;
RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF00419; BAB02958.1;
SQ SEQUENCE 841 AA; 93216 MW; 7DBEE72618EB8B97 CRC64;

```

Query Match 10.54; Score 425; DB 2; Length 841;
 Best Local Similarity 23.04; Pred. No. 2.3e-17; Indels 240; Gaps 35;
 Matches 200; Conservative 122; Mismatches 308; Indels 240; Gaps 35;

```

QY 56 SSSSPATATTTATLVSLPESIDRPERHVPIDIFYQVLAQTHFLTDGIRRAFEARVSK 115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 81 SRTSSLAASTST-----TELPLYCYQLIGVSEQAEKDEVKSVINLKRT 124
QY 116 PPGFSGDALLISRRQILQACETLSNPSRRRRYNGSLDDEATVITDVPMKVKVGLC 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 125 DAEEGYMEAAARQQLMDVDRKL--LPESYAGNLKEKIAKPSRLIPMAMLPGLC 181
QY 176 VLQEGSTETVLRVGEALLKERLPKSFQDVVLMALFLVDSRDAMALDPDPFITGYEF 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 LLOBVGQKLVLDIGRAALRNDSKPYINDIFLSMALBACAIKAPFENVKVS--QGFEA 239
QY 236 VBEALKLQEE-GASSIAPDLRAQIDETLEETPRVYLELGLPLDDVYAKLNGISGV 294
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 LABAQFLSKVTLGKLA--LTLQIESLEELAPCTLLDGLPRTPENERRRGAIAL 297
QY 295 RNILMSVGGGASALVGLTRK-----PMNEAFLEMTAAEQVDLF-----VATP 339
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 298 RELLRQ-----GLSVEASCOIOWPCTLSQAISSLATLETIVDLPLWMDLAI TRK 346
QY 340 SNIPASFE-----YEVALLVAQAFIGKPEHLLODADKQFOLOAKVAMEIP 390
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 347 NKSLSHNRVVIDPNCPTMULGHIAVGFSG-----KQNETINKATTICEGL 396
QY 391 AMLDYTRNNWEIDFGLERGLCALLIGKDEEC-RMWLGIDSE-----DSQY 434
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 397 A-----SEGVDLKEFEAFCSFLKQLSATGPTCWLQASSEALEKQLKLESNSDSAV 449
QY 435 RNPAIVEFVLENSNRDDNDLPGLCKLETWLAGVPPRRDRK-----DKKF 482
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 450 RNS-----LIGKESRST-----ATPSLEAMLESVLANPDRPGSPSLANFFRAKKT 499
QY 483 KLGDYDDPWLVS-----LEFEVVOGSPLAAMATMARIGAEHVAKASAMQALQVPS 536
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 500 PENKKGSPSINMHKTNQRLSTQGVNS-----QHL----- 532
QY 537 RYDRNSABPKDQERYFVSVDVPGNNVGRDBPGVFIAEAVRSENETNDVAIRAGVSE 596
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 533 -YTAVEQLPPTDQSPVSAK--NNDE-----TSASMPSVOLKRN--LGVHK 574
QY 597 SSV-DETYEMSVADMLKEASVKIL-----AAGVAILGISLF-----SQ 634
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 575 NKIWDEN--LSQSSILGRVSVALGCTVFFSLKULSGIRSGRLQSNPISVSARPHSESD 631
QY 635 KYFLKSSSFQKDMVSMESDVATIGSVRA----- 665
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 632 SPIMKTESGNFRKQNLDSVNRNGI--VGNIKVLIDMLKMGHBPALYLKSSGQSATSLS 689
QY 666 -DDSEALPR-MDARTAENIVSKWKIKSLAFQPDRIEMLEPVLDGRM-----LK 713
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 690 HSASBLHKPMDREBAELVQWENYKAENLGPTHQVYSLSVELDSMLVQVESIFLCIM 749
QY 714 IWTDRAEATQOLGVY-DYTLKLSTDSVTVSAD---GTRALVEATLEESACLSDLVHP 769
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 750 QW-QTLAQTAEAKSCYWRFLHLLEVLQAHIFEDGAGENAETLEALBEALVDSQPK 808
QY 770 NNATDVRYTTRRYEVFMSKSG-WKITEGSV 798
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 809 -NAKYSTYKIRYILKKQBDGLMKFCQSDI 837

```

RESULT 13

```

Q8DKU7 PRELIMINARY; PRT; 673 AA.
ID Q8DKU7
AC Q8DKU7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T10758 protein.
GN OrderedlocusNames=cl0758;
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.

```


OX NCBI_TaxID=32046;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BP-1;
 RX MEDLINE=2225144; PubMed=12240834;
 RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
 RA Watanabe A., Iritaguchi M., Kawashima K., Kimura T., Kishida Y.,
 RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
 RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the thermophilic cyanobacterium
 Thermosynechococcus elongatus BP-1.";
 RL DNA Res. 9:123-130(2002).
 DR EMBL; AP005371; BAC08309.1;
 DR InterPro; IPR001623; DnaJ_N.
 DR InterPro; IPR008941; TPR-like.
 DR Pfam; PF00226; DnaJ_1.
 DR SMART; SM00271; DnaJ_1.
 DR PROSITE; PS0076; DnaJ_2; 1.
 DR Complete proteome.
 SO SEQUENCE 673 AA; 75741 MW; 3427C6E46BCB83A CRC64;
 Query Match 10.4%; Score 423; DB 2; Length 673;
 Best Local Similarity 25.4%; Pred. No. 2.2e-17;
 Matches 191; Conservative 105; Mismatches 307; Indels 150; Gaps 28;
 OY 86 IPIDFYVGLAQTHFLTDGIRRAFEARVSKRPGFSDALISRQILQAACETLSNPRS 145
 DB 3 IPIDFYVGLAQTHFLTDGIRRAFEARVSKRPGFSDALISRQILQAACETLSNPRS 62
 OY 146 RREYNE-----GLLDEEATVTVDPMDKVGALCVLOEGSTETVLKRGAL 194
 DB 63 RDADVRRCTVDPDDLQAQDPDATTTHIEISDEQSGALLLVELGNVAVVGLGAF 122
 OY 195 KERLPKSFQ-----DVLVMAFLVDSRDAMALDPDFTTGEFEVBAK 243
 DB 123 KKVY---FEERRPYTSPAAVADITLVATLALTLGRBEM--QRQSYSAASQLRAGLQVL 177
 OY 244 QEEGASLAPDLRAQIDETLEITPRVYLIELGLPLGDDVAAKRLNGSVRNILMVG 303
 DB 178 QR---VNLPELQGFQTELRNRPYRIELALPLSD--SANRQIGILKQVLSERG 232
 OY 304 -GGASALVGLTRB--KFMNEAFLEMTAEQVDFVATPSNIPAESFEVEVALVAQ 359
 DB 233 IEGGDDRSGLTVEDPKFLIQLRSHLTVAHQELF-----BRESRRBAVATLYAH 285
 OY 360 AFIGKKEHLID----ADKQFQLOAKWAMEIPAMLIDTRNNWEIDFGLRGCLAL 414
 DB 286 ALVARGVHEIQPSYICRAKDLILQOL-----LPHQ-----DVLIELASCLLL 326
 OY 415 IGKVDCEGRMVLGLDSESOYRNPAIVEFVLENSNRDNDLPGCLKLETTVLAVGPRF 474
 DB 327 LGGQTE-----ALALHSDQDPTL-DPIRRHAG-EGAGDRPLGLYYTTQGLTEITPAF 379
 OY 475 RDTCKKKFKLDYDDPMVLSTYERVEVVOGSPPLAAATMARIGABHVKASAMQALQVF 534
 DB 380 RDLGETFVALBAVADANVOYTLALSEDSIAPRPPTA----- 419
 OY 535 PSRYTDRNSAEPKDVQGVTFVSDVGNVNGGEGVFIABAVRPSNFETNDVAIRAGV 594
 DB 420 -----SALPEYIRPTV-AVPP-----PLSFTAEVL-PLDQSRILGGGLSASA 459
 OY 595 SESSVDETTYEMSVADMLKEASVK-----ILAAVVA-IGLISLSQKFLKSS 642
 DB 460 FTSPATATGTSMQPSPRKRSPNRCAQKQTFMFGAGVVLGAL--AKYWPDAKT 517
 OY 643 SPORKDVSMSDVATIGSVRADSEAL--PRMDARTAEIVSKMOKISLAFGPDHR 699
 DB 518 AEAAPPPVTAAPVATPTPTPOPTTIAITITPEM-----ADRLHTYQQLKAQGLGPPPE 573
 OY 700 IEMLPEVLDGRMLKIWDRAAETAGLGVYDYTLKLSVDSVTVS--ADGTRALVE-----A 754
 DB 574 VDKLTITLAEPELSRMRSAQGLKSEGSYVYTLKLNLEKVEKRLQSRDRAVLAIRVEDA 633

OY 755 TLRESACSLDVHPENNATDVRYTTRVEFWS 787
 DB 634 RPYEQGTL-----RTDI-STSDPYRYIYT 656
 RESULT 14
 ID Q7VOH1 PRELIMINARY; PRT; 702 AA.
 AC Q7VOH1;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Hypothetical protein.
 GN OrderdlocusNames=PMW1287;
 OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
 OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococaceae;
 OC Prochlorococcus.
 OX NCBI_TaxID=59919;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;
 RA Kocap G., Larimer F.W., Lamerdin J.E., Malfatti S., Chait P.,
 RA Ahlgren N.A., Arellano A., Coleman M., Hauser L., Hees W.R.,
 RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,
 RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
 RA Webb E.A., Zinser E.R., Chisholm S.W.;
 RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
 RT niche differentiation."
 RL Nature 424:1042-1047(2003).
 DR EMBL; BX572093; CAE19746.1;
 DR InterPro; IPR001623; DnaJ_N.
 DR Pfam; PF00226; DnaJ_1.
 DR SMART; SM00271; DnaJ_1.
 DR Complete proteome; Hypothetical protein.
 SO SEQUENCE 702 AA; 79938 MW; 325A8F3E74E558F CRC64;
 Query Match 8.5%; Score 344.5; DB 2; Length 702;
 Best Local Similarity 20.5%; Pred. No. 1.5e-12;
 Matches 161; Conservative 155; Mismatches 305; Indels 163; Gaps 31;
 OY 86 IPIDFYVGLAQTHFLTDGIRRAFEARVSKRPGFSDALISRQILQAACETLSNPRS 145
 DB 3 IPIDFYVGLAQTHFLTDGIRRAFEARVSKRPGFSDALISRQILQAACETLSNPRS 62
 OY 146 RREYNEGLDDEEATVTVDPMDKVGALCVLOEGSTETVLKRGALK---ERLPKS 201
 DB 63 RRYVENILINGASG--LDSLSNREVAGLILWESSGSKFAFKTRKRLAOPOTPALGSS 119
 OY 202 FKQDVVLMALFLDVRDA--MALDPDFTTGEFEFEALKLLQEEGASLAPDLRAQI 259
 DB 120 READLTLAAL---TSRDAIIEODORSYSNAADFLQESIQLOLRMGKLG--ELKTL 172
 OY 260 DETLEITPRVYVELLGLPLGDDVAAKRLNGSVRNILMVGSGGASALVGLTRB--K 317
 DB 173 EBDLVSLTPRIIDLSRLNDVDSHK--GLMTEHLIIKRG-----LEGKKSRYND 225
 OY 318 PMN---EAFLR-----MTAAEQVDFVATPSNIPAESFEVEVALVAQAFIKKPHL 368
 DB 226 FLNQEESESFQOIKPFLTVQDOIIFLELQKRGSSBA--GFLAFLSLTIAIGFARRPAK 283
 OY 369 LQADKQFQLOAKWAME-----IPAMLVDTRNNWEIDFGLRGCLALIGKVDCEGRM 424
 DB 284 LFEARKLTKLNLISGLDSMPILICLIDLILAD-----VEQSSARFLSSDDEKLRDW 333
 OY 425 LGLDSEDSQYRNPAIVEFVLENSNRDNDLPGCLKLETTVLAVGPPRDTCKKKFKL 484
 DB 334 L-----NNYRGEKLEAICIFCKRWLENDVAGRIDDKLEIDL 371
 OY 485 GDYDDPMVLSTYERVEVVOGSPPLAAATMARIGABHVKASAMQALQKVPSTRYDRNSA 544
 DB 372 DSWFEDEREIOEFIOIE-----KKSNTYFKSGPON-----KPIFOAOSLSLDSS 416
 OY 545 EPKDVQGVTFVSD--PVGNVGRDG-----EPGFIABAVRPSNFETNDVAIRAGVESS 598

Db	417	TGPDLSNDNFEBGRJLPLPGCVARBDQGEVIEENIYDTEIK-NKSTIEFYKAIK-TAE--	472
Qy	599	VDETTVMSVADMLKE-----ASVKILAAVAGI.I-----SLFSQRYFLK	639
Db	473	-----LKFPAGSALBNRYIFNKSSYLTYLAFILIPAFIGVGVFANNLKKPVQEKREID	527
Qy	640	SSSFQRKDMV--SSMESD-----VATISVAPADSE-----ALPRDADTAEN	661
Db	528	NLSLSTENKNVVEYEGILNODKKKVLNDNSKIILSDNKEVIFSGEIKTASPSLE--KIEH	585
Qy	682	IVSKQKIKSLAFGPDHRIEMPEVLVDGRMLKIWTBPAETAQLGVVDYDTLLKLSVDSV	741
Db	586	LINTVLYNKSCKRLAGKEINLSKIYQDDDLIDRLKEREID-IÖKG-I--YKININAIENI	641
Qy	742	TVSADGTRALVEATLEESACSLDLVHPE-----NNADIVRTYTRYREYEVMSKGV	791
Db	642	-----VLQTQASRISVSDLKYESEKILKIDGELINETFTPLPKVYKILIGFSNNSW	693
Qy	792	KITE 795	
Db	694	KLVD 697	

RESULT 15			
07U557			
ID	07U557	PRELIMINARY;	PRT: 653 AA.
AC	07U557;		
DT	01-OCT-2003 (Tremblurel, 25, Created)		
DT	01-OCT-2003 (Tremblurel, 25, Last sequence update)		
DT	01-OCT-2003 (Tremblurel, 25, Last annotation update)		
DE	Hypotheetical protein.		
GN	Ordered locus names=SYNM1619;		
OS	Synechococcus sp. (Strain WH8102).		
OC	Bacteria; Cyanobacteria; Chroococcales; Synechococcus.		
OX	NCBI_TaxID=84588;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=22825697; PubMed=12917641; DOI=10.1038/nature01943;		
RA	Palenik B., Brahamsha B., Laitner F.W., Land M.L., Hauser L.,		
RA	Chen P., Lamerdin J.E., Regala W., Allen E.B., McCarren J.,		
RA	Puilean I.T., Dufresne A., Patentsky F., Webb E.A., Waterbury J.;		
RT	"The genome of a motile marine Synechococcus.";		
RL	Nature 424:1037-1042(2003).		
DR	EMBL: BX569693; CAE08134.1; -		
DR	InterPro: IPR001623; DnaE_N		
QW	Complete proteome; Hypotheetical protein.		
QW	SEQUENCE 653 AA; 70997 MW; 4bDf7926F67BEC37 CRC64;		

Query Match	8.44;	Score 340;	DB 2;	Length 653;
-------------	-------	------------	-------	-------------

Best Local Similarity 21.5%; Pred. No. 2.6e-12;
Matches 163; Conservative 120; Mismatches 317; Indels 156; Gaps 24;

```

QY      86  IPIIDYOVYAGTHTFLTDGIRAFARVSKPPQFGSDDLIRSDIOLAACTSLNPRS 145
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      3  LPIDHRLFGVSPSADPASILKRLQRRSSPBDGCTHGGLOKQALLHRSADLTDPE 62
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      146 RREYNELGD----DEBQTVITDVPWDKYPGALCYLQSGEETIVLRGEBALLKERLP- 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      63 RADYEAALLISTSATHPNERTVGLDLAASSSEVAGILILMEGALLAEVQLARQGIQPPQAPA 122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      200 --KSPKQDVLVYMALAFIVSRDGMALDPPDFITGYEPFEBALKLQEBGASLAPLRA 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      123 LQSGREADITTLAALACRDAADEQ--QQRVYSAQMLRDGIELQRMKG---LPDQA 177
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      258 QIDETLEETIPRYVELLGLPLGDDYAAKRLNGLSGVRNILMSVG-----GASALV 310
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      178 RLQGLDLDLPLRYVLDLSRDLSD--ADARQGISILDLQVVRDGGIDPGLDSETPYAM 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      311 GGLTEKPKNNKFLKRTAAEOVDLFLFA--TPSNIPASFEFVYVALALVAQAATIGKKPHL 366
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      236 GQADPESFPQQLRRFLTVGEOVDLFGFGWPAEGSIEGCLAVF---ALAAAGYSRRKPEF 291
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY      369  LODADKQEQOLQOQKVMAMEIPAMLYOTRNNWEIDFGERGICALLIGVDCRMVLGLD 428
Db      292  LEQAREOLQQRIVASGLDMPPL-----LGLDILLGNVAASLHF----- 330
QY      429  SEDSQYNPAIVEFVLENSNRDNDLPELCKLLETLAGVFPFRPDTKDKKFKLGDIY 488
Db      331  ---SAIRDEELISWLAEH-----PCGHILAQCEYCRWMLERDVLPGYRDVDAAGVDIDAMF 383
QY      489  DDPMWLSTLERVEVQGSPLAAAMAIIGAEHVYASAMQOLQKVPFSHYTRBNSAEPD 548
Db      384  ADRDQOAVDRID-RQSARLGSAAVTGTAG-----LSAPSDAASS 423
QY      549  VQETVFSVDYPPGNNGVRGDEBGFVIAEAVRPSENEETNDYAIRAGVSESSVDETVEMSV 608
Db      424  PHEALDDDH-----PAEEAPSSD-----P 444
QY      609  ADMLKEASVKLLAGVAIGLISLFSQKFLPKSSSF-----QRKMVSSMES--DVA 658
Db      445  ANORLSNLRWMLASLTVYGLVAALAAAVMLRRETRAPVVLQEPDRQDVEPRKPSQDSA 504
QY      659  TI-----GSYRADADEALPRM-----DARTENTVYSKWQIKSLAFGPDHRIEMPEVL 707
Db      505  TLKPOLIIQPEASEVAQOLQPLLSAPBDQAOLRMVQGHLSKQAQL--QGQPSQLPVYA 562
QY      708  DGRMLKIWTDRPAETTAQGLVYDYTLKLSDVSVTVSADGTR-----ALVEATLEES 759
Db      563  RQRLIID-QVDR--ERSKAVVAGTTTVVAKSVYSLDVSVRQOPRIEIQAOVAASDYSTRDS 619
QY      760  ACLSGLVHPENNADVRYTYTTRYRYEYFWSKSGKIT 794
Db      620  GTIVDRIVP-----GSLTITTYIIIGRDGDQDKLT 647

```

Search completed: June 10, 2005, 01:40:34
Job time : 104 secs

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2005, 01:37:44 ; Search time 475 Seconds
(without alignments)
646.423 Million cell updates/sec

Title: US-10-600-070-2

Perfect score: 4063
Sequence: 1 MEALSHVIGIGLSPFQCRPLP.....YEVWMSKMGKITEGVTLAS 801

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1710399 seqs, 383334425 residues

Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*

3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubppaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubppaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/1/pubppaa/US10D_PUBCOMB.pep:*

17: /cgn2_6/ptodata/1/pubppaa/US10E_PUBCOMB.pep:*

18: /cgn2_6/ptodata/1/pubppaa/US10F_PUBCOMB.pep:*

19: /cgn2_6/ptodata/1/pubppaa/US11A_PUBCOMB.pep:*

20: /cgn2_6/ptodata/1/pubppaa/US11_NEW_PUB.pep:*

21: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*

22: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4063	100.0	801	US-10-600-070-2	Sequence 2, Appl
2	4051	99.7	801	US-10-600-070-127	Sequence 127, Appl
3	4051	99.7	801	US-10-600-070-129	Sequence 129, Appl
4	4051	99.7	801	US-10-600-070-131	Sequence 131, Appl
5	4051	99.7	801	US-10-739-930-5771	Sequence 5771, App
6	2926	92.0	578	US-10-600-070-123	Sequence 123, App
7	1775.5	43.7	760	US-10-600-070-125	Sequence 125, App
8	1646	40.5	324	US-10-600-070-126	Sequence 126, App
9	1332	32.8	525	US-10-600-070-122	Sequence 122, App
10	1134	27.9	344	US-10-424-599-271849	Sequence 271849, App
11	973	23.9	364	US-10-425-115-266516	Sequence 266516, App

12	972	23.9	480	US-10-437-963-172416	Sequence 172416, App
13	856	21.1	416	US-10-425-115-242115	Sequence 242115, App
14	524.5	12.9	768	US-10-600-070-167	Sequence 167, App
15	499	12.3	204	US-10-424-599-177901	Sequence 177901, App
16	498.5	12.3	789	US-10-600-070-194	Sequence 194, App
17	497	12.2	798	US-10-600-070-164	Sequence 164, App
18	497	12.2	798	US-10-600-070-165	Sequence 165, App
19	482	11.9	709	US-10-437-963-172415	Sequence 172415, App
20	474.5	11.7	1157	US-10-437-963-155799	Sequence 155799, App
21	442	10.9	631	US-10-600-070-5	Sequence 5, Appl
22	442	10.9	631	US-10-600-070-162	Sequence 162, App
23	435.5	10.7	191	US-10-767-701-35633	Sequence 35633, A
24	434.5	10.7	624	US-10-600-070-116	Sequence 116, App
25	431	10.6	819	US-10-600-070-171	Sequence 171, App
26	431	10.6	819	US-10-600-070-173	Sequence 173, App
27	429	10.6	714	US-10-600-070-169	Sequence 169, App
28	429	10.6	714	US-10-600-070-170	Sequence 170, App
29	425.5	10.5	652	US-10-600-070-115	Sequence 115, App
30	423	10.4	673	US-10-600-070-192	Sequence 192, App
31	413	10.2	836	US-10-600-070-190	Sequence 190, App
32	409.5	10.1	205	US-10-767-701-41038	Sequence 41038, A
33	401	9.9	716	US-10-600-070-160	Sequence 160, App
34	366	9.0	99	US-10-600-070-124	Sequence 124, App
35	359	8.8	566	US-10-600-070-119	Sequence 119, App
36	344.5	8.5	702	US-10-600-070-156	Sequence 156, App
37	339	8.3	66	US-10-600-070-100	Sequence 100, App
38	328.5	8.1	573	US-10-600-070-120	Sequence 120, App
39	327	8.0	357	US-10-600-070-117	Sequence 117, App
40	320	7.9	661	US-10-600-070-158	Sequence 158, App
41	317.5	7.8	515	US-10-600-070-121	Sequence 121, App
42	308.5	7.6	524	US-10-600-070-118	Sequence 118, App
43	288	7.1	164	US-10-437-963-110670	Sequence 110670, App
44	263	6.5	66	US-10-600-070-99	Sequence 99, Appl
45	257	6.3	16	US-10-600-070-98	Sequence 98, Appl

ALIGNMENTS

RESULT 1

US-10-600-070-2

Sequence 2, Application US/10600070

Publication No. US20040139500A1

GENERAL INFORMATION:

APPLICANT: Osteeryoung, Katherine W.

APPLICANT: Vitsha, Stanislaw

APPLICANT: Koksharova, Olga A.

APPLICANT: Gao, Hongbo

TITLE OF INVENTION: Placid Division and Related Genes and Proteins, and Methods of

TITLE OF INVENTION: Use

FILE REFERENCE: MSU-08153

CURRENT APPLICATION NUMBER: US/10/600,070

CURRENT FILING DATE: 2003-06-20

NUMBER OF SEQ ID NOS: 206

SOFTWARE: PatentIn version 3.2

SEQ ID NO 2

LENGTH: 801

TYPE: PRT

ORGANISM: Arabidopsis thaliana

US-10-600-070-2

Query Match 100.0%; Score 4063; DB 16; Length 801;

Beet Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;

Matches 801; Conservative 0; Indels 0;

Qy 1 MEALSHVIGIGLSPFQCRPLPATTKLRSNTSTTICSSAKMDRLSDNFSTSSSSSS 60

Db 1 MEALSHVIGIGLSPFQCRPLPATTKLRSNTSTTICSSAKMDRLSDNFSTSSSSSS 60

Qy 61 PAATTTATVSLPPSIDRPRRHVPIIDRYQVGAQTHFLTDGIRAFKRVSKPQFG 120

Db 61 FATATTATVSLPPSIDRPRRHVPIIDRYQVGAQTHFLTDGIRAFKRVSKPQFG 120

QY 121 FSDDALISRQOIIQAACETLSNPRSREYNEGILLDBEATVITDVPMDKVPALCVI.QEG 180
| | | | |
DB 121 FSDDALISRQOIIQAACETLSNPRSREYNEGILLDBEATVITDVPMDKVPALCVI.QEG 180
| | | | |
QY 181 GETEIVLRVGEALLKRLPKSPFKODVVLVWALAFLDVSRDAMALDPDPFITGGEFVEAL 240
| | | | |
DB 181 GETEIVLRVGEALLKRLPKSPFKODVVLVWALAFLDVSRDAMALDPDPFITGGEFVEAL 240
| | | | |
QY 241 KLIQEBGASSLAPDLRAQIDETLEETPRVYLBGLPLGDDVAAKRLNGLSGVRNLTMS 300
| | | | |
DB 241 KLIQEBGASSLAPDLRAQIDETLEETPRVYLBGLPLGDDVAAKRLNGLSGVRNLTMS 300
| | | | |
QY 301 VGGGASALVGLTRREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVEVALALVAQA 360
| | | | |
DB 301 VGGGASALVGLTRREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVEVALALVAQA 360
| | | | |
QY 361 FIGKKPHLLQDADKQFQOLQOAKVMAEIPAMLVDTNNMEIDFGLERGLCALLIGKVD 420
| | | | |
DB 361 FIGKKPHLLQDADKQFQOLQOAKVMAEIPAMLVDTNNMEIDFGLERGLCALLIGKVD 420
| | | | |
QY 421 CRMWLGIDSEDSQYRNPAIYEFVLENSNRDNDLPGCLKLETLWLAGVFPFRFRTKDX 480
| | | | |
DB 421 CRMWLGIDSEDSQYRNPAIYEFVLENSNRDNDLPGCLKLETLWLAGVFPFRFRTKDX 480
| | | | |
QY 481 KFKLGDIYDDPMVLSYERVEVVOGSPLAAMATMARI GAEHVKAAMQALOKVPSRYTD 540
| | | | |
DB 481 KFKLGDIYDDPMVLSYERVEVVOGSPLAAMATMARI GAEHVKAAMQALOKVPSRYTD 540
| | | | |
QY 541 RNAABPDVQETVPSVDPVGNNGRDEPGVFIAEAVRPSNEFTNDYAIRAGVSESSVD 600
| | | | |
DB 541 RNAABPDVQETVPSVDPVGNNGRDEPGVFIAEAVRPSNEFTNDYAIRAGVSESSVD 600
| | | | |
QY 601 ETTVEMSVADMKEASVKILAAVAIGLISLFSQKFLKSSSFQORDMWSMESDVATI 660
| | | | |
DB 601 ETTVEMSVADMKEASVKILAAVAIGLISLFSQKFLKSSSFQORDMWSMESDVATI 660
| | | | |
QY 661 GSVRADSEALPRMDARTAEIVSKWOKISLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
| | | | |
DB 661 GSVRADSEALPRMDARTAEIVSKWOKISLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
| | | | |
QY 721 ETTQGLGVYDTLLKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
| | | | |
DB 721 ETTQGLGVYDTLLKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
| | | | |
QY 781 RYEVFWSKSGMKITTEGSVLAS 801
| | | | |
DB 781 RYEVFWSKSGMKITTEGSVLAS 801
| | | | |

RESULT 2

US-10-600-070-127
; Sequence 127, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: OsteYoung, Katherine W.
; APPLICANT: Vittha, Stanislaw
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Honggo
; TITLE OF INVENTION: Placid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 127
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-600-070-127

Query Match 99.7%; Score 4051; DB 16; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;

Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MEALSHVIGLSFQOLCRLEPPATTKLRSNHTSTTICSAKMADRLLSDNFTSDSSSS 60
| | | | |
DB 1 MEALSHVIGLSFQOLCRLEPPATTKLRSNHTSTTICSAKMADRLLSDNFTSDSSSS 60
| | | | |
QY 61 PATRTTATVLSLPISIDRPERHVPIDFYQVLAGQTHFLTGIRAFARVSKPQFG 120
| | | | |
DB 61 PATRTTATVLSLPISIDRPERHVPIDFYQVLAGQTHFLTGIRAFARVSKPQFG 120
| | | | |
QY 121 FSDDALISRQOIIQAACETLSNPRSREYNEGILLDBEATVITDVPMDKVPALCVI.QEG 180
| | | | |
DB 121 FSDDALISRQOIIQAACETLSNPRSREYNEGILLDBEATVITDVPMDKVPALCVI.QEG 180
| | | | |
QY 181 GETEIVLRVGEALLKRLPKSPFKODVVLVWALAFLDVSRDAMALDPDPFITGGEFVEAL 240
| | | | |
DB 181 GETEIVLRVGEALLKRLPKSPFKODVVLVWALAFLDVSRDAMALDPDPFITGGEFVEAL 240
| | | | |
QY 241 KLIQEBGASSLAPDLRAQIDETLEETPRVYLBGLPLGDDVAAKRLNGLSGVRNLTMS 300
| | | | |
DB 241 KLIQEBGASSLAPDLRAQIDETLEETPRVYLBGLPLGDDVAAKRLNGLSGVRNLTMS 300
| | | | |
QY 301 VGGGASALVGLTRREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVEVALALVAQA 360
| | | | |
DB 301 VGGGASALVGLTRREKFMNEAFLRMTAAEQVDLFVATPSNIPAESFEVEVALALVAQA 360
| | | | |
QY 361 FIGKKPHLLQDADKQFQOLQOAKVMAEIPAMLVDTNNMEIDFGLERGLCALLIGKVD 420
| | | | |
DB 361 FIGKKPHLLQDADKQFQOLQOAKVMAEIPAMLVDTNNMEIDFGLERGLCALLIGKVD 420
| | | | |
QY 421 CRMWLGIDSEDSQYRNPAIYEFVLENSNRDNDLPGCLKLETLWLAGVFPFRFRTKDX 480
| | | | |
DB 421 CRMWLGIDSEDSQYRNPAIYEFVLENSNRDNDLPGCLKLETLWLAGVFPFRFRTKDX 480
| | | | |
QY 481 KFKLGDIYDDPMVLSYERVEVVOGSPLAAMATMARI GAEHVKAAMQALOKVPSRYTD 540
| | | | |
DB 481 KFKLGDIYDDPMVLSYERVEVVOGSPLAAMATMARI GAEHVKAAMQALOKVPSRYTD 540
| | | | |
QY 541 RNAABPDVQETVPSVDPVGNNGRDEPGVFIAEAVRPSNEFTNDYAIRAGVSESSVD 600
| | | | |
DB 541 RNAABPDVQETVPSVDPVGNNGRDEPGVFIAEAVRPSNEFTNDYAIRAGVSESSVD 600
| | | | |
QY 601 ETTVEMSVADMKEASVKILAAVAIGLISLFSQKFLKSSSFQORDMWSMESDVATI 660
| | | | |
DB 601 ETTVEMSVADMKEASVKILAAVAIGLISLFSQKFLKSSSFQORDMWSMESDVATI 660
| | | | |
QY 661 GSVRADSEALPRMDARTAEIVSKWOKISLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
| | | | |
DB 661 GSVRADSEALPRMDARTAEIVSKWOKISLAFGPDHRIEMLPVLDGRMLKIWTDRAA 720
| | | | |
QY 721 ETTQGLGVYDTLLKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
| | | | |
DB 721 ETTQGLGVYDTLLKLSVDSVTSADGTRALVEATLEESACLSDLVHPENNATDVRTYTT 780
| | | | |
QY 781 RYEVFWSKSGMKITTEGSVLAS 801
| | | | |
DB 781 RYEVFWSKSGMKITTEGSVLAS 801
| | | | |

RESULT 3

US-10-600-070-129
; Sequence 129, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: OsteYoung, Katherine W.
; APPLICANT: Vittha, Stanislaw
; APPLICANT: Koksharova, Olga A.
; APPLICANT: Gao, Honggo
; TITLE OF INVENTION: Placid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20

NUMBER OF SEQ ID NOS: 206
SOFTWARE: Patentin version 3.2
SEQ ID NO 129
LENGTH: 801
TYPE: PR
ORGANISM: Arabidopsis thaliana
US-10-600-070-129

Query Match 99.7%; Score 4051; DB 16; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 MEALSHVIGIGLSPFQLCRLPPATTKLRSHNTSTTICSASAKMADRLSDNFSTDSSSSS 60
DB 1 MEALSHVIGIGLSPFQLCRLPPATTKLRSHNTSTTICSASAKMADRLSDNFSTDSSSSS 60
QY 61 FATATTATLVSPSIDRPERHVPIDFYQVLAQTHFLTDGIRAFARVSKPPQFG 120
DB 61 FATATTATLVSPSIDRPERHVPIDFYQVLAQTHFLTDGIRAFARVSKPPQFG 120
QY 121 FSDDALISRQIIQAACETISNPSRREYNEGLDDEAATVITDVPMDKVPALCVLQEG 180
DB 121 FSDDALISRQIIQAACETISNPSRREYNEGLDDEAATVITDVPMDKVPALCVLQEG 180
QY 121 FSDDALISRQIIQAACETISNPSRREYNEGLDDEAATVITDVPMDKVPALCVLQEG 180
DB 121 FSDDALISRQIIQAACETISNPSRREYNEGLDDEAATVITDVPMDKVPALCVLQEG 180
QY 181 GETEIVLRVEBALIKERLPKSFQDVVLAALFLDVSADAMALDPDFITGYEFVEBAL 240
DB 181 GETEIVLRVEBALIKERLPKSFQDVVLAALFLDVSADAMALDPDFITGYEFVEBAL 240
QY 181 GETEIVLRVEBALIKERLPKSFQDVVLAALFLDVSADAMALDPDFITGYEFVEBAL 240
DB 181 GETEIVLRVEBALIKERLPKSFQDVVLAALFLDVSADAMALDPDFITGYEFVEBAL 240
QY 241 KLIQEBGASSIAPDLRAQIDETLEETPRVYLBELGLPLGDDVAKKLNGLSGVRNIIIMS 300
DB 241 KLIQEBGASSIAPDLRAQIDETLEETPRVYLBELGLPLGDDVAKKLNGLSGVRNIIIMS 300
QY 301 VGGGASALVGLTREKFMNEAFIRMTAABQVDLFVATPSNIPAESEVEVALALVAQA 360
DB 301 VGGGASALVGLTREKFMNEAFIRMTAABQVDLFVATPSNIPAESEVEVALALVAQA 360
QY 301 VGGGASALVGLTREKFMNEAFIRMTAABQVDLFVATPSNIPAESEVEVALALVAQA 360
DB 301 VGGGASALVGLTREKFMNEAFIRMTAABQVDLFVATPSNIPAESEVEVALALVAQA 360
QY 361 FIGKPHLLDADKQFQOLQAQKNAMEIPAMLVDTNNMEIDFGLRGICALLIGKVD 420
DB 361 FIGKPHLLDADKQFQOLQAQKNAMEIPAMLVDTNNMEIDFGLRGICALLIGKVD 420
QY 421 CRMWLGIDSEDSQYRNPAIYEFVLNSNRDNDLPGICKLETFWLAGVFPFRDTRDK 480
DB 421 CRMWLGIDSEDSQYRNPAIYEFVLNSNRDNDLPGICKLETFWLAGVFPFRDTRDK 480
QY 421 CRMWLGIDSEDSQYRNPAIYEFVLNSNRDNDLPGICKLETFWLAGVFPFRDTRDK 480
DB 421 CRMWLGIDSEDSQYRNPAIYEFVLNSNRDNDLPGICKLETFWLAGVFPFRDTRDK 480
QY 481 KFKLGDYDDPMVLSYERVEVVGSSPLAAATMARI GAETHVKSAMQALQKVPSTRYTD 540
DB 481 KFKLGDYDDPMVLSYERVEVVGSSPLAAATMARI GAETHVKSAMQALQKVPSTRYTD 540
QY 481 KFKLGDYDDPMVLSYERVEVVGSSPLAAATMARI GAETHVKSAMQALQKVPSTRYTD 540
DB 481 KFKLGDYDDPMVLSYERVEVVGSSPLAAATMARI GAETHVKSAMQALQKVPSTRYTD 540
QY 541 RNSAEPDQVETVSVDPVGNNGRDEPGVFI EAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPDQVETVSVDPVGNNGRDEPGVFI EAVRPSNFETNDYAIRAGVSESSVD 600
QY 541 RNSAEPDQVETVSVDPVGNNGRDEPGVFI EAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPDQVETVSVDPVGNNGRDEPGVFI EAVRPSNFETNDYAIRAGVSESSVD 600
QY 601 ETTVEMSVADMKEASVKIILAAVAIGLISLFSQKYLKSSSSFORQDMVSMSDVATI 660
DB 601 ETTVEMSVADMKEASVKIILAAVAIGLISLFSQKYLKSSSSFORQDMVSMSDVATI 660
QY 601 ETTVEMSVADMKEASVKIILAAVAIGLISLFSQKYLKSSSSFORQDMVSMSDVATI 660
DB 601 ETTVEMSVADMKEASVKIILAAVAIGLISLFSQKYLKSSSSFORQDMVSMSDVATI 660
QY 661 GSVRADSEALPRMDARTAEINIVSKWQIKSLAGPDRHIREMPEVLDGRMLKIWDRAA 720
DB 661 GSVRADSEALPRMDARTAEINIVSKWQIKSLAGPDRHIREMPEVLDGRMLKIWDRAA 720
QY 721 ETTAQLGLVYDITLKLKSVDSVTVSADGTRALVEATLESACLSDLVHPENNATDVRITYTT 780
DB 721 ETTAQLGLVYDITLKLKSVDSVTVSADGTRALVEATLESACLSDLVHPENNATDVRITYTT 780
QY 781 RYEVFMKSGMKITGEGSVLAS 801
DB 781 RYEVFMKSGMKITGEGSVLAS 801
```

RESULT 4
US-10-600-070-131
Sequence 131, Application US/10600070
Publication No. US20040139500A1

GENERAL INFORMATION:
APPLICANT: Oseeryoung, Katherine W.
APPLICANT: Vilha, Stanislaw
APPLICANT: Koksharova, Olga A.
APPLICANT: Gao, Hongo
TITLE OF INVENTION: Placid Division and Related Genes and Proteins, and Methods of
FILE REFERENCE: NSU-08153
CURRENT APPLICATION NUMBER: US/10/600,070
NUMBER OF SEQ ID NOS: 206
SOFTWARE: Patentin version 3.2
SEQ ID NO 131
LENGTH: 801
TYPE: PR
ORGANISM: Arabidopsis thaliana
US-10-600-070-131

Query Match 99.7%; Score 4051; DB 16; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 MEALSHVIGIGLSPFQLCRLPPATTKLRSHNTSTTICSASAKMADRLSDNFSTDSSSSS 60
DB 1 MEALSHVIGIGLSPFQLCRLPPATTKLRSHNTSTTICSASAKMADRLSDNFSTDSSSSS 60
QY 61 FATATTATLVSPSIDRPERHVPIDFYQVLAQTHFLTDGIRAFARVSKPPQFG 120
DB 61 FATATTATLVSPSIDRPERHVPIDFYQVLAQTHFLTDGIRAFARVSKPPQFG 120
QY 121 FSDDALISRQIIQAACETISNPSRREYNEGLDDEAATVITDVPMDKVPALCVLQEG 180
DB 121 FSDDALISRQIIQAACETISNPSRREYNEGLDDEAATVITDVPMDKVPALCVLQEG 180
QY 121 FSDDALISRQIIQAACETISNPSRREYNEGLDDEAATVITDVPMDKVPALCVLQEG 180
DB 121 FSDDALISRQIIQAACETISNPSRREYNEGLDDEAATVITDVPMDKVPALCVLQEG 180
QY 181 GETEIVLRVEBALIKERLPKSFQDVVLAALFLDVSADAMALDPDFITGYEFVEBAL 240
DB 181 GETEIVLRVEBALIKERLPKSFQDVVLAALFLDVSADAMALDPDFITGYEFVEBAL 240
QY 181 GETEIVLRVEBALIKERLPKSFQDVVLAALFLDVSADAMALDPDFITGYEFVEBAL 240
DB 181 GETEIVLRVEBALIKERLPKSFQDVVLAALFLDVSADAMALDPDFITGYEFVEBAL 240
QY 241 KLIQEBGASSIAPDLRAQIDETLEETPRVYLBELGLPLGDDVAKKLNGLSGVRNIIIMS 300
DB 241 KLIQEBGASSIAPDLRAQIDETLEETPRVYLBELGLPLGDDVAKKLNGLSGVRNIIIMS 300
QY 241 KLIQEBGASSIAPDLRAQIDETLEETPRVYLBELGLPLGDDVAKKLNGLSGVRNIIIMS 300
DB 241 KLIQEBGASSIAPDLRAQIDETLEETPRVYLBELGLPLGDDVAKKLNGLSGVRNIIIMS 300
QY 301 VGGGASALVGLTREKFMNEAFIRMTAABQVDLFVATPSNIPAESEVEVALALVAQA 360
DB 301 VGGGASALVGLTREKFMNEAFIRMTAABQVDLFVATPSNIPAESEVEVALALVAQA 360
QY 301 VGGGASALVGLTREKFMNEAFIRMTAABQVDLFVATPSNIPAESEVEVALALVAQA 360
DB 301 VGGGASALVGLTREKFMNEAFIRMTAABQVDLFVATPSNIPAESEVEVALALVAQA 360
QY 361 FIGKPHLLDADKQFQOLQAQKNAMEIPAMLVDTNNMEIDFGLRGICALLIGKVD 420
DB 361 FIGKPHLLDADKQFQOLQAQKNAMEIPAMLVDTNNMEIDFGLRGICALLIGKVD 420
QY 361 FIGKPHLLDADKQFQOLQAQKNAMEIPAMLVDTNNMEIDFGLRGICALLIGKVD 420
DB 361 FIGKPHLLDADKQFQOLQAQKNAMEIPAMLVDTNNMEIDFGLRGICALLIGKVD 420
QY 421 CRMWLGIDSEDSQYRNPAIYEFVLNSNRDNDLPGICKLETFWLAGVFPFRDTRDK 480
DB 421 CRMWLGIDSEDSQYRNPAIYEFVLNSNRDNDLPGICKLETFWLAGVFPFRDTRDK 480
QY 421 CRMWLGIDSEDSQYRNPAIYEFVLNSNRDNDLPGICKLETFWLAGVFPFRDTRDK 480
DB 421 CRMWLGIDSEDSQYRNPAIYEFVLNSNRDNDLPGICKLETFWLAGVFPFRDTRDK 480
QY 481 KFKLGDYDDPMVLSYERVEVVGSSPLAAATMARI GAETHVKSAMQALQKVPSTRYTD 540
DB 481 KFKLGDYDDPMVLSYERVEVVGSSPLAAATMARI GAETHVKSAMQALQKVPSTRYTD 540
QY 481 KFKLGDYDDPMVLSYERVEVVGSSPLAAATMARI GAETHVKSAMQALQKVPSTRYTD 540
DB 481 KFKLGDYDDPMVLSYERVEVVGSSPLAAATMARI GAETHVKSAMQALQKVPSTRYTD 540
QY 541 RNSAEPDQVETVSVDPVGNNGRDEPGVFI EAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPDQVETVSVDPVGNNGRDEPGVFI EAVRPSNFETNDYAIRAGVSESSVD 600
QY 541 RNSAEPDQVETVSVDPVGNNGRDEPGVFI EAVRPSNFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPDQVETVSVDPVGNNGRDEPGVFI EAVRPSNFETNDYAIRAGVSESSVD 600
QY 601 ETTVEMSVADMKEASVKIILAAVAIGLISLFSQKYLKSSSSFORQDMVSMSDVATI 660
DB 601 ETTVEMSVADMKEASVKIILAAVAIGLISLFSQKYLKSSSSFORQDMVSMSDVATI 660
QY 601 ETTVEMSVADMKEASVKIILAAVAIGLISLFSQKYLKSSSSFORQDMVSMSDVATI 660
DB 601 ETTVEMSVADMKEASVKIILAAVAIGLISLFSQKYLKSSSSFORQDMVSMSDVATI 660
QY 661 GSVRADSEALPRMDARTAEINIVSKWQIKSLAGPDRHIREMPEVLDGRMLKIWDRAA 720
DB 661 GSVRADSEALPRMDARTAEINIVSKWQIKSLAGPDRHIREMPEVLDGRMLKIWDRAA 720
QY 721 ETTAQLGLVYDITLKLKSVDSVTVSADGTRALVEATLESACLSDLVHPENNATDVRITYTT 780
DB 721 ETTAQLGLVYDITLKLKSVDSVTVSADGTRALVEATLESACLSDLVHPENNATDVRITYTT 780
```

QY 781 RYEFWMSGKMTTEGSVLAS 801
DB 781 RYEFWMSGKMTTEGSVLAS 801

RESULT 5
US-10-739-930-5771
; Sequence 5771, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739, 930
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 5771
; LENGTH: 801
; TYPE: PR1
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C13643_1.p
US-10-739-930-5771

Query Match 99.7%; Score 4051; DB 16; Length 801;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 799; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MEALSHVIGLSPPQLCRLPPATTKLRSHNTSTTICSAKMDRLISDFNTSDSSSS 60
DB 1 MEALSHVIGLSPPQLCRLPPATTKLRSHNTSTTICSAKMDRLISDFNTSDSSSS 60
QY 121 FSDALISRRQIIQAACETSNPRSRREYEGLLDDEEATVITDVPMDKPGALCVLOEG 180
DB 121 FSDALISRRQIIQAACETSNPRSRREYEGLLDDEEATVITDVPMDKPGALCVLOEG 180
QY 121 FSDALISRRQIIQAACETSNPRSRREYEGLLDDEEATVITDVPMDKPGALCVLOEG 180
DB 121 FSDALISRRQIIQAACETSNPRSRREYEGLLDDEEATVITDVPMDKPGALCVLOEG 180
QY 181 GETEIVLRVEGALLKERLPKSFQDQVVLVMAFLDVSMDAMALDPDFITGYEFVEAL 240
DB 181 GETEIVLRVEGALLKERLPKSFQDQVVLVMAFLDVSMDAMALDPDFITGYEFVEAL 240
QY 241 KLIQEBGASLAPDLRAQIDETLEETPRVYELGLPLGDDVYAAKRLNGLSGRNLTMS 300
DB 241 KLIQEBGASLAPDLRAQIDETLEETPRVYELGLPLGDDVYAAKRLNGLSGRNLTMS 300
QY 301 VGGGASALVGGLTREKFNMEAFLRMTAAEQVDLFVATPSNIPAESFEYEVALLVAQA 360
DB 301 VGGGASALVGGLTREKFNMEAFLRMTAAEQVDLFVATPSNIPAESFEYEVALLVAQA 360
QY 361 FICKKPHLLQDADKQFOQLQAAKVMAMEIPAMLVDTNNMEIDFGLERGLCALLIGKVD 420
DB 361 FICKKPHLLQDADKQFOQLQAAKVMAMEIPAMLVDTNNMEIDFGLERGLCALLIGKVD 420
QY 421 CRWMLGLDSRDSQYRNPATVEFYLENSNRDNDLPGCLCKLETWLAGVFPFRDTRDK 480
DB 421 CRWMLGLDSRDSQYRNPATVEFYLENSNRDNDLPGCLCKLETWLAGVFPFRDTRDK 480
QY 481 KFKLGDYDDPMTLSYLERVEVVOGSPLAATAATMARI GAHVKASAMQALOKVFPSTYTD 540
DB 481 KFKLGDYDDPMTLSYLERVEVVOGSPLAATAATMARI GAHVKASAMQALOKVFPSTYTD 540
QY 541 RNSAEPKDVQETVPSYDVGNNVGRDGPVFLAEAVRPSSENFETNDYAIRAGVSESSVD 600
DB 541 RNSAEPKDVQETVPSYDVGNNVGRDGPVFLAEAVRPSSENFETNDYAIRAGVSESSVD 600
QY 601 ETTVEHSVADMLKEASVKILAAAGVAILGLISLFSQKFLKSSSFQKDMVSSMESDVATI 660
DB 601 ETTVEHSVADMLKEASVKILAAAGVAILGLISLFSQKFLKSSSFQKDMVSSMESDVATI 660

QY 661 GSVRADDSSEALPMDARTANIVSKWQIKSLAFGPDHRIEMLPEVLGDMKIMTDRAA 720
DB 661 GSVRADDSSEALPMDARTANIVSKWQIKSLAFGPDHRIEMLPEVLGDMKIMTDRAA 720
QY 721 ETRQGLVYDVTYLLKLSVDSVTVSADGTRALVEATLEESACSLDVHPENNATDVRTYTT 780
DB 721 ETRQGLVYDVTYLLKLSVDSVTVSADGTRALVEATLEESACSLDVHPENNATDVRTYTT 780
QY 781 RYEFWMSGKMTTEGSVLAS 801
DB 781 RYEFWMSGKMTTEGSVLAS 801

RESULT 6
US-10-600-070-123
; Sequence 123, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Osteeryoung, Katherine W.
; APPLICANT: Vilcha, Stanislav
; APPLICANT: Koksharova, Olga A.
; TITLE OF INVENTION: Placetic Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 123
; LENGTH: 578
; TYPE: PR1
; ORGANISM: Arabidopsis thaliana
US-10-600-070-123

Query Match 72.0%; Score 2926; DB 16; Length 578;
Best Local Similarity 99.8%; Pred. No. 2.8e-237;
Matches 577; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 79 RPRRHVPIPIIDFYQVGAQTHPLTDSIRRAFEARVSKPPQFSPSDALISRRQIIQAAC 138
DB 1 RPRRHVPIPIIDFYQVGAQTHPLTDSIRRAFEARVSKPPQFSPSDALISRRQIIQAAC 60
QY 139 TSNPSRSREYEGLLDDEEATVITDVPMDKPGALCVLOEGGETEIVLRVGBALLKERL 198
DB 139 TSNPSRSREYEGLLDDEEATVITDVPMDKPGALCVLOEGGETEIVLRVGBALLKERL 120
QY 61 TSNPSRSREYEGLLDDEEATVITDVPMDKPGALCVLOEGGETEIVLRVGBALLKERL 120
DB 61 TSNPSRSREYEGLLDDEEATVITDVPMDKPGALCVLOEGGETEIVLRVGBALLKERL 120
QY 199 PKSFQDQVVLVMAFLDVSMDAMALDPDFITGYEFVEALKLIQEBGASLAPDLRAQ 258
DB 199 PKSFQDQVVLVMAFLDVSMDAMALDPDFITGYEFVEALKLIQEBGASLAPDLRAQ 180
QY 121 PKSFQDQVVLVMAFLDVSMDAMALDPDFITGYEFVEALKLIQEBGASLAPDLRAQ 180
DB 121 PKSFQDQVVLVMAFLDVSMDAMALDPDFITGYEFVEALKLIQEBGASLAPDLRAQ 180
QY 259 IDETLEETPRVYELGLPLGDDVYAAKRLNGLSGRNLTMSVGGGASALVGGLTREK 318
DB 259 IDETLEETPRVYELGLPLGDDVYAAKRLNGLSGRNLTMSVGGGASALVGGLTREK 240
QY 319 MNEAFLRMTAAEQVDLFVATPSNIPAESFEYEVALLVAQAFTGKPKPHLLQDADKQFOQ 378
DB 319 MNEAFLRMTAAEQVDLFVATPSNIPAESFEYEVALLVAQAFTGKPKPHLLQDADKQFOQ 300
QY 241 MNEAFLRMTAAEQVDLFVATPSNIPAESFEYEVALLVAQAFTGKPKPHLLQDADKQFOQ 300
DB 241 MNEAFLRMTAAEQVDLFVATPSNIPAESFEYEVALLVAQAFTGKPKPHLLQDADKQFOQ 300
QY 379 LQQAATMAMEIPAMLVDTNNMEIDFGLERGLCALLIGYDRCRWMLGLDSRDSQYRNP 438
DB 379 LQQAATMAMEIPAMLVDTNNMEIDFGLERGLCALLIGYDRCRWMLGLDSRDSQYRNP 360
QY 301 LQQAATMAMEIPAMLVDTNNMEIDFGLERGLCALLIGYDRCRWMLGLDSRDSQYRNP 360
DB 301 LQQAATMAMEIPAMLVDTNNMEIDFGLERGLCALLIGYDRCRWMLGLDSRDSQYRNP 360
QY 439 IVEFYLENSNRDNDLPGCLCKLETWLAGVFPFRDTRDKKFKLGDYDDPMTLSYTB 498
DB 439 IVEFYLENSNRDNDLPGCLCKLETWLAGVFPFRDTRDKKFKLGDYDDPMTLSYTB 420
QY 361 IVEFYLENSNRDNDLPGCLCKLETWLAGVFPFRDTRDKKFKLGDYDDPMTLSYTB 420
DB 361 IVEFYLENSNRDNDLPGCLCKLETWLAGVFPFRDTRDKKFKLGDYDDPMTLSYTB 420
QY 499 RVEVVOGSPPLAAATAATMARI GAHVKASAMQALOKVFPSTYTDNRNSAEPKDVQETVSV 558
DB 499 RVEVVOGSPPLAAATAATMARI GAHVKASAMQALOKVFPSTYTDNRNSAEPKDVQETVSV 480
QY 421 RAEVVOGSPPLAAATAATMARI GAHVKASAMQALOKVFPSTYTDNRNSAEPKDVQETVSV 480
DB 421 RAEVVOGSPPLAAATAATMARI GAHVKASAMQALOKVFPSTYTDNRNSAEPKDVQETVSV 480
QY 559 VGNNVGRDGPVFLAEAVRPSSENFETNDYAIRAGVSESSVD ETTVEHSVADMLKEASVK 618
DB 559 VGNNVGRDGPVFLAEAVRPSSENFETNDYAIRAGVSESSVD ETTVEHSVADMLKEASVK 618

Db	Qy	Db
481	ILAGVAILGLISLFSQKTFLEKSSSFQKDMVSMESD	656
541	ILAGVAILGLISLFSQKTFLEKSSSFQKDMVSMESD	578

RESULT 7

```

US-10-600-070-125
: Sequence 125, Application US/10600070
: Publication No. US20040139500A1
: GENERAL INFORMATION:
: APPLICANT: Oseeryoung, Katherine W.
: APPLICANT: Vitka, Stanislav
: APPLICANT: Kokeharova, Olga A.
: APPLICANT: Gao, Hongbo
: TITLE OF INVENTION: Plasmid Division and Related Genes and Proteins, and Methods of
: TITLE OF INVENTION: Use
: FILE REFERENCE: MSU-08153
: CURRENT APPLICATION NUMBER: US/10/600,070
: CURRENT FILING DATE: 2003-06-20
: NUMBER OF SEQ ID NOS: 206
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 125
: LENGTH: 760
: TYPE: PR1
: ORGANISM: Oryza sativa
US-10-600-070-125

```

Query Match	43.78;	Score 1775.5;	DB 16;	Length 760;
Query Match	43.78;	Score 1775.5;	DB 16;	Length 760;

Best Local Similarity 47.9%; Pred. No. 4.1e-140;
Matches 390; Conservative 119; Mismatches 213; Indels 93; Gaps 16.

QY 12 SPFQLCRLPATTKLRSHTSTTIC-SASKWADRLSDNF-----TSDSSSSSFATAT 65

D5 14 APFAPSLPRPRPRPRPHPSACRASRWABRLFADFHLPLTAAPSDPPSPAPAPA 73

[illegible][illegible][illegible]

04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

[illegible]

246 EGASSLPDLRAQIDETLBEITPRVYLRLGLPIGDDYAKRI^{NGI}SGVRNTIWSVGGGG 30

Db 245 DGASNLAPDLLSOIDETLEEITPRCVLELLSLPIDTEHHKROEGLOGARNILWSVGRGG 30

306 ASALVGGLTREKEMNEAFILRMTAAEQVDLFVATPSNIPAESFEVYEVALALVAQAFIGKK 36

Db 305 IATVGGGFSRAFMNEAFLRMTSIEQMDFFSKTPNSIPPEWFEIYNVALAHVAQAIISKR 36

366 PHLLQDADKQFQOLQQAkvMAMEI PAMLYDTRNNWEIDFGLERGLCALLIGKVDECRMWL 42

Db 365 PQFIMMADDLFEQLQKFNIGS---HYAYDN---EMDLALERAFCSLVGVDSKCRMWL 41

426 GUDSBDSDQIKNPALVEFVIBNSN-KUDNDDLPGLCKLBIBWAGVFFPRFRDINDNKFNL 48

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

[illegible][illegible]

```

Db      537 OLDSAMENT-----KDG-PCGYT-----ENFDO----- 55
      ||::|      ::|      ||:

```

```

OY 598 SYDETYEMSAVDMLKESVYTLAAGVAILGISLESOXKFEKSSSFQKMDWMSNEO- 656
Db 560 -----ENAPHDOSRNALAKITSAGALFALLAIVIGAK-----LPKRPLSATRSEH 605
OY 657 --VATIGSVRADDSSEAL-----PRMDARTAEINVSXKQKIKSLAFGPDHRIEMPEYL 707
Db 606 GSVAIVANSVDSITDOPDALDEDPVHIIRPMQKLAEDIVRKMQSISKXALGPEHSAVSLQEBYL 665
OY 708 DGRMLKIWTDBAAEYAOUGLYVDYITLLKLSVDSVSVASADGTALVEATIBESACSLDVH 767
Db 666 DGNMLKWITDAAABIERHGWEMFEBYLSVTIDSTIISLDGRRAVEATIDEAGOLTDVTE 725
OY 768 PENNAATDVRITTYTRIEVFSMK--SGWKIREGSYLAS 801
Db 726 PRNNDSDYTKTYTRIEYEMAFSKLGWKITIEGANLKS 760

```

RESULT 8

```

US-10-600-070-206
; Sequence 206, Application US/10600070
; Publication No. US20040139500A1
; GENERAL INFORMATION:
; APPLICANT: Oseeryoung, Katherine W.
; APPLICANT: Vitsha, Stanislav
; APPLICANT: Kokeharova, Olga A.
; APPLICANT: Gao, Hongbo
; TITLE OF INVENTION: Placitid Division and Related Genes and Proteins, and Methods of
; TITLE OF INVENTION: Use
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 206
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-600-070-206

```

Query Match	Score	DB	Length
40.5%	1646	16	324

```

Best Local Similarity 100.0%; P-vec. NO. 8.0e-130;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 MEALSHVGLSPFQLCRLPPATIKLRKSHNISTITCSASKWADRLLSDFNFISDSSSSS 60

DD I MEALSHVGIUGLSFFYLUCKUEFPAI I LUKRSHN I S I I LCBASHWADUKLESDFNF I SDSSSSSS DV

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

[illegible]

Q. 181 C E T T I Y D V G E A I K E B I P K S E K O D V Y T Y M A I A F I D V S B D A M A I D B P D E T T G V E E V E R A I. 240

Db 181 GETETIVLRVGEALTKERLPKSEKQDVTLMALAEIDVSRDAMALDPDPETITGYEEVEEAL 240

QY 241 KLIQEEGASLAPDLRAQIDETLEEITPRVLELLGLPLGDDYAAKRLNGLSGVRILWS 300

Db 241 KLTQEGASSLPDLRAQIDETLEEITPRVYLELGLPLGDDYAAKRLNGLSGVRNLWS 300

QY 301 VGGGASALVGLTRKFMNEAFL 3 24

Db 301 VGGGASALVGGLTRKFMNEAFL 324

RESULT 9

US-10-000-070 122
; Sequence 122, Application US/10600070

Public

Publication No. US20040139500A1

```

; GENERAL INFORMATION:
; APPLICANT: Osteoryoung, Katherine W.
; APPLICANT: Viltha, Stanislaw
; APPLICANT: Kosharova, Olga A.
; APPLICANT: Gao, Hongbo
; TITLE OF INVENTION: Placitid Division and Related Genes and Proteins, and Methods of
; FILE REFERENCE: Use
; FILE REFERENCE: MSU-08153
; CURRENT APPLICATION NUMBER: US/10/600,070
; CURRENT FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 122
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-600-070-122

Query Match      32.8%; Score 1332; DB 16; Length 525;
Best Local Similarity 48.8%; Pred. No. 5,4e-103;
Matches 287; Conservative 88; Mismatches 137; Indels 76; Gaps 11;

QY 81 ERHVPIPIDYOVYLGACTHFLTDGIRAFARVSKPQGFSDDALSRQIIQAACETL 140
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 3 ERSLPIQVDYKVLGAPHPFLGDSIRPAFARIAPQYIGSTDALGRRQMLQIADTL 62
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 141 SNPSRREYNEGILDDDEATVITDVPWDKYPGALCVLQEGGETEIVLRGEALIKERLPK 200
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 63 MNQNSTQYVRALSENREALITMDIAMDK-----EAGBALAVLVTEQGLLDRPK 113
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 201 SFQODVYVLMALAFDVSRDAMALDPDFTTGYEFVEBALKILQEGSALPDLRAQID 260
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 114 RFQODVYVLMALAYVDLSRDAMASPDPVIGCEVEERALKILOEDASNIADPDLISQID 173
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 261 ETELEETPRVYVLELGLPLGDDVAAKRLNGISGVNIIWSVGSGGASALVGLTREKFN 320
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 174 ETELEETPRVYVLELGLPLDTEHHKROBGLQGRNIIWSVGSGGASALVGLTREKFN 223
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 321 EAFRLMTAAEQVDLFLVTPSNIPAESFEVEVALALVAQAFIGKKPHLLQDADKQFQQLQ 380
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 234 EAFRLMTSIEQOMFPFKTPIPSIRPEWEIYNVALAHVAQAIISRRPQFIMMADLPFQQLQ 293
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 381 QAKVMAEIPAMLYDTRNNEIDPGLERGLCALLIGKVECRMWLGIDSDSOTRNPAIV 440
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 294 KFNIGS-----HYAYDN-----EMDLALERAFCSLIVGVDSKCRMWLGIDNESSPYRDKIL 345
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 441 EPLERSEN-RDDDDDLPGLCLEMTLAGVPRPRPTKKPKFLGYYDDPWLSTYLER 499
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 346 EPLVTNSSIIEENDLPGLCLEMTLWLFVFPSPSRDTRGMQFRLGYYDDPEVLSTYLER 405
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 500 VEVVQSSPLAAATMTARIGAE-----HYKASAMQALQKVPF-SRYTDRNSAEPKQVQET 552
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 406 MEGGASHLAAAAAKLGAQATRACTVSNIAQANKYFPLIEQUDRAMENT----- 460
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 553 VESVDVGNVNGDEBPVFIAEAVRPSNPETNDVIAIRAGVSESSVDTEVKEVADML 612
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 461 -----KDG-FGGYL-----ENPFDQ-----ENAPAHDS 481
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 613 KEASVKTILAGVAIGLISLFSQKYFLKSSSPQRKQNVSSMESGVANT 660
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 482 RNHALKITSIGALPALAVIGAKY-----LPRKPLSAIRSHSGSV 522
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 10
US-10-424-599-271849
; Sequence 271849, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

```

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 271849
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_87499C.1.pcp
US-10-424-599-271849

Query Match      27.9%; Score 1134; DB 15; Length 344;
Best Local Similarity 66.4%; Pred. No. 1.3e-86;
Matches 223; Conservative 48; Mismatches 53; Indels 12; Gaps 3;

QY 16 LCRLPATY-----KLRSS--HNTSTTICSAKADRLSPNTSDSSSFATATY 66
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 12 LCTPHPTTYTPFKPNKLLRSSLSRGAASLSATSKWAERLIADPFLGDMAAS--TSTS 68
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 67 TATVSLPSIDRPERHVPPIPIFYOVYLGACTHFLTDGIRAFARVSKPQGFSDDAL 126
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 69 TISPSVPRPDPPEHYVSIPLDYRILGAEHPFLDGIIRAVEAKFSKRPQYAFSDAL 128
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 127 ISRRQIIQAACETLSNPSRREYNEGILDDDEATVITDVPWDKYPGALCVLQEGGETEIV 186
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 129 ISRRQIIQAACETLADPPTSREYVQSLVDDEFAIITQPIPFDKYPGALCVLQEGGETEIV 188
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 187 LRVGEALIKERLPKSKRQDVYVLMALAFDVSRDAMALDPDFTTGYEFVEBALKILQOE 246
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 189 LRTGGILRRRLPKTERKQDVYVLMALAFDVSRDAMALSPDPIAACEMLEERALKILOER 248
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 247 GASSLAPDLRAQIDETLEETPRVYVLELGLPLGDDVAAKRLNGISGVNIIWSVGSGGA 306
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 249 GATSLAPDLRAQIDETLEETPRVYVLELGLPLDDEHRAAREBGILGVNIIWAVGGGA 308
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 307 SALVGLTREKFNNEAFLRMTAAEQVDLFPVATPSNT 342
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 309 ATTAGFTREDPFMNEAFLHMTAAEQVELFPVATPSNT 344
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

RESULT 11
US-10-425-115-266516
; Sequence 266516, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222) B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 266516
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_174665C.1.pcp
US-10-425-115-266516

Query Match      23.9%; Score 973; DB 16; Length 364;
Best Local Similarity 54.8%; Pred. No. 5,4e-73;
Matches 198; Conservative 54; Mismatches 79; Indels 30; Gaps 4;

QY 12 SPQOLCRLPATYKLRSHNT-----STTICSAKMAADRLSDENF-----TSDSS 57
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 14 APFAFPLP-----LKRSHRPPPPGPGSTCRASRWADRLFADPHLLPMAADPPMASSSS 68
   |||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```


Db 204 DDLKDS----LEKLAQGSVAGDAIHDSRNDAL-----KIIISAGTLTALFAVLGLKCLPRN 254
QY 641 SSSFORKDWVSMESDVATIGSVRADDSBA-----LPRMDARTANIYSKKMKIKSLA 693
Db 255 KS-----LPALRGEGSVAVVDSIDGPPADDEPLEIPMDAKLAIDIYRRMOSIKSKA 307
QY 694 FGPDHRIEMLEPVLDGEMLKIMTDRAAETAGLGVDTYTLKLKSVDSVTVSADGTRALVE 753
Db 308 LGEHETVLTALQVLTGGMMLKMTDRABERIRHGMFWFETALSGVITDSITVSVGRRAAVE 367
QY 754 ATIEESACSLDVLHPENNATDVRTYTTTRYEVFWSK--SGWKITSGSVLAS 801
Db 368 ATIEEVGRLLTDVDPKNDADYDTKTYARVEMTYSRPAQWRTGEGAVLKS 416

RESULT 14

US-10-600-070-167
/ Sequence 167, Application US/10600070
/ Publication No. US20040139500A1
/ GENERAL INFORMATION:
/ APPLICANT: Oetereyong, Katherine W.
/ APPLICANT: Vitha, Stanislaw
/ APPLICANT: Koksharova, Olga A.
/ APPLICANT: Gao, Hongbo
/ TITLE OF INVENTION: Placitid Division and Related Genes and Proteins, and Methods of
/ FILE REFERENCE: MSU-08153
/ CURRENT APPLICATION NUMBER: US/10/600,070
/ CURRENT FILING DATE: 2003-06-20
/ NUMBER OF SEQ ID NOS: 206
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 167
/ LENGTH: 768
/ TYPE: PRN
/ ORGANISM: Nostoc punctiforme
US-10-600-070-167

Query Match 12.9%; Score 524.5; DB 16; Length 768;
Best Local Similarity 24.9%; Pred. No. 1.1e-34;
Matches 212; Conservative 128; Mismatches 286; Indels 227; Gaps 28;

QY 84 VPIPIPIFYQVLAGAOTHEFLTDCIRAFARVSKPPQFGSDALISRRQIILOACETISNP 143
Db 1 VRIPLDYRILGLPLAASEQOLROAVSDRIVOLPRREYSQALISSRKLIEBAYVVLSDP 60
QY 144 RSRREYNEGL-----DDEBAYVIT-----DVPMDKVGKALCVLQ 178
Db 61 KQSTYDQLYLAAYDPDNLAAAQAQENRTSTKSGSDTQSLGIEITDDELGALLILQ 120
QY 179 EGGETEIVLAVGERALT--KERLPKSPKQ-----DVVLVMAIAFLDVSRD- 220
Db 121 ELGSEYELVNLGRPYLVNKSATSSRSKNNLADBEIYESAHPVVLVTVALACLELREQ 180
QY 221 -----AMALDPDPTIGYEFVEBALKULOEBGASSIAPDLRAQIDITLLEITPRV 271
Db 181 MOOCHYENNAISLE-----TGOS-----LVREG---LFSSIOAEIOADLYLRPYRI 225
QY 272 LELLGLPLGDVYAKRLNGISGRNLTMSVGG--GGSALVGGILTR--KPMNAEFLRMT 327
Db 226 LELLALP--OEKTAERSQGLLELQNLLEDRGGIDGTNNDESGNLIDPLFLFIQOLRNHLT 283
QY 328 AAEQVLDLVATPSNIPAESFEVYEVVALVAQAFTGKPHILODADKPOOLOQAKVMAM 387
Db 284 VAEQHQLFBAQSKR--SSAVATYLAVALYLAIRGPAQRPALINGAQMVLGRQ----- 337
QY 388 EIRPAMLYTRNNKMEIDPGLERGLCALLIGVYDSCRMMLGLDSEDSQYRNPALVEFVLNS 447
Db 338 -----DVHLEQSLCALILGQTBETATRLBELSOF---YE---ALAFIREKS 376
QY 448 NRDDNDLPLGLCLLETWLAGVVPFRPRDKKKFKGDIYDDPMVLSTYLERVEVVGSP 507
Db 377 -QDSPDLPLGLCLYAEQMLQHEVFPFRFLDANQOAFLLKDYFANQOVAYILE----- 426

QY 508 LAAATMARIAGAEHVKASAMQALQKVPSPRYTDRNSAEPK-----DVQETVFSV 556
Db 427 --ALPTDAQTTNEMAVINPQYFPQAKAKNTHPHANNSTKTSASFNHARVPMDLPEI----- 480
QY 557 DPGNVNNGDGEFGVPIAEVVRPSENFETNDYAIRAGVSSSDVETTVENSVDMLK--- 613
Db 481 -----PTK--ETSEYDNFSPMWSGSIKSEVPAERMSRGT 516
QY 614 -----EASVKILAG----- 623
Db 517 NQHLNLSAASASAGNHQKRRKPTSPASRERIPDNRPHSRRRRRRTFANTIEGKTRLV 576
QY 624 --VAIGLISI-----FSQYFLKSSSFQKDWVSMESDVATIGSVRADDS 669
Db 577 WRVFISILVFWPLATTTFGWLKNLFFPQSPDPDLQFLVQINQPLPIPDNRKRESE 636
QY 670 ALPRMDARTENIVSKQKIKSLAFGPDHRIEMLEPVLDGEMLKIMTDRAAETAGLGV 729
Db 637 EGPLVTNAE--AEVYIHWLSTKAALGPNHENNLLEQILTGSALSQWR--LIAQONKLDNRY 694
QY 730 ---DYTLKLKSVDSVTVSADGTRALVEATLEESACSLDVLHPENNATDVRTYTTREVEFW 786
Db 695 RKPDHSLIKIESVAKIGLFPAD--RAAYEATVKEVTLQYENNQFQSSND--KLARVIDLIR 750
QY 787 SKSGWKITGCVL 799
Db 751 ERGKWRIGQTSVV 763

RESULT 15

US-10-424-599-177901
/ Sequence 177901, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J
/ APPLICANT: Zhou Yihua
/ APPLICANT: Zhou Yihua
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53223) B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 177901
/ LENGTH: 204
/ TYPE: PRN
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_131660C.1.pep
US-10-424-599-177901

Query Match 12.3%; Score 499; DB 15; Length 204;
Best Local Similarity 51.2%; Pred. No. 1.9e-33;
Matches 104; Conservative 39; Mismatches 58; Indels 2; Gaps 2;

QY 601 ETTVEWSVADMLKEASVKILAGVAIGLISLFSQKYF-LKSSSFQKDWVSMESDVAT 659
Db 2 ETRKGFPIIEIRIGHASVQIMCAGVIGLTVLGLKFLPRNNGSPIARKMTGSMAMVDGIN 61
QY 660 TGSVRADDS--EALPRMDARTENIVSKQKIKSLAFGPDHRIEMLEPVLDGEMLKIMTDR 718
Db 62 LGSIGDEBEKEVQLPKMDARVAEALVWKQSVKSEAGCPDHCLGRLEHVLIDGEMLKIMTDR 121
QY 719 AAEATQGLVYDVTLLKLSVDSVTVSADGTRALVEATLEESACSLDVLHPENNATDVRTY 778
Db 122 AAEIARSGSYDITLDDINLDSVTISONGRRRAVEITLKESTHNLAVGHQPHDASNSRTY 181
QY 779 TTRYEVFWSKSGWKITGCVLAS 801
Db 182 TTRYEVFWSKSGWKITGCVLAS 204

✓ Fri Jun 10 09:57:55 2005

us-10-600-070-2.rapb

Page 9

Search completed: June 10, 2005, 01:50:42
Job time : 477 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 9, 2005, 15:02:56 ; Search time 1188 Seconds
(without alignments)
12476.017 Million cell updates/sec

Title: US-10-600-070-3

Perfect score: 3667

Sequence: 1 tgcctgcataaggaagat.....ctataacataagagctacaa 3667

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_g881.*
9: gb_g882.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	725.2	19.8	919	3	CNS09YUH
2	436.2	11.9	741	3	CNS09ZSS
3	434.6	11.9	741	3	CNS09YNM
4	402.2	11.0	561	1	A1998415
5	363.6	9.9	785	8	BZ437564
6	359.4	9.8	434	5	BP620404
7	336.4	9.2	897	6	CD573714
8	315.8	8.6	534	5	B0834167
9	311.8	8.5	320	1	AV830764
10	307.2	8.4	320	1	BZ765587
11	305.6	8.3	954	5	CG961431
12	295.6	8.1	631	5	BU046755
13	293.4	8.0	832	7	CO079829
14	265.4	7.2	751	9	CR486014
15	254	6.9	849	7	CO075595
16	252.6	6.9	769	7	CO075595
17	252	6.9	813	7	CO071968
18	239.6	6.5	423	1	AV812946
19	237.2	6.5	703	7	CNS18842
20	212.6	5.8	521	7	CNS14655
21	212.6	5.8	593	7	CNS904734
22	210	5.7	660	4	BI268376
23	208.8	5.7	660	4	BI268376
24	194	5.3	2307	9	CL965374

25	193.2	5.3	510	7	CF603268
26	166.4	4.5	542	5	B0586205
27	161.8	4.4	819	7	CV241483
28	161.8	4.4	821	7	CK090561
29	161.8	4.4	922	7	CV264112
30	161	4.4	545	5	B0410206
31	161	4.4	692	9	CL723297
32	158	4.3	608	2	BE472035
33	157.8	4.3	722	7	CO117046
34	157.4	4.3	321	7	AL936312
35	156.2	4.3	947	9	CG088802
36	155.6	4.2	764	6	CD903230
37	153	4.2	723	7	CO079828
38	152.6	4.2	574	2	AM696905
39	152.4	4.2	611	6	CA173502
40	148.8	4.1	866	6	CD573715
41	148.2	4.0	683	7	CO076192
42	147.8	4.0	258	6	CD479481
43	147.8	4.0	914	8	BZ722703
44	147.6	4.0	476	5	B0835742
45	147.4	4.0	652	9	CL584270

ALIGNMENTS

RESULT 1	CNS09YUH	919 bp	mrna	linear	HTC 04-FEB-2004
LOCUS	Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSJTS892C08 of Adult vegetative tissue of strain col-0 of Arabidopsis thaliana (thale cress).				
DEFINITION	CNS09YUH	919 bp	mrna	linear	HTC 04-FEB-2004
ACCESSION	EX841670.1	GI:42406830			
VERSION	HTC; GSI: cDNA.				
KEYWORDS	Arabidopsis thaliana (thale cress)				
SOURCE	Arabidopsis thaliana				
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
REFERENCE	Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C., Menard,M., Cruaud,C., Queiller,F., Searpelli,C., Schachter,V., Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.				
AUTHORS	Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences: A Combined Approach to Evaluate and Improve Arabidopsis Genome Annotation				
TITLE	Unpublished				
JOURNAL	Genoscope.				
REFERENCE	Direct Submission				
AUTHORS	Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage ; BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr				
JOURNAL	Web : www.genoscope.cns.fr				
COMMENT	The sequences are based on single pass reads. Life Technologies (a division of Invitrogen) members carried out full-length libraries construction ; Temple G. Genoscope members carried out sequencing and annotation ; Castelli V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M. UNG INRA : Clepet C., Caboche M. Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences) . 5 prime and 3 prime are assembled with Pirap. http://www.genoscope.cns.fr/externe/sequences/Banque_projet_EF/Full length http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.				
FEATURES	Location/Qualifiers				
SOURCE	1..919				
	/organism="Arabidopsis thaliana"				
	/mol_type="mrna"				
	/strain="Col-0"				
	/db_xref="taxon:3702"				

/clone="GSLTILS892C08"
/tissue_type="Adult vegetative tissue"
/plasmid="pCMVSPORT_6"
1. 919
/gene="At5g42480"

ORIGIN

Query Match 19.8%; Score 725.2; DB 3; Length 919;
Best Local Similarity 84.2%; Pred. No. 5.7e-176;
Matches 916; Conservative 0; Mismatches 3; Indels 169; Gaps 2;

QY 469 AAAGACACTCCCATGGAAGCTCTGAGTCAGCTCGGCAATGGTCTCTCCCATTCOAATTA 528
DB 1 AAAGACACTCCCATGGAAGCTCTGAGTCAGCTCGGCAATGGTCTCTCCCATTCOAATTA 60

QY 529 TGGCGATTACACCGGGGAGAGCAAAAGCTCCGAGTGGCCCAACACCTCTACAAATTC 588
DB 61 TGGCGATTACACCGGGGAGAGCAAAAGCTCCGAGTGGCCCAACACCTCTACAAATTC 120

QY 589 TGCTCCGCGCAAGCAATGGGCGGACCGTCTTCTCCGACTTCAATTTCACTCCGATTCC 648
DB 121 TGCTCCGCGCAAGCAATGGGCGGACCGTCTTCTCCGACTTCAATTTCACTCCGATTCC 180

QY 649 TCCTCTCTCTCTCTGCGCACCGCCACCAACCGCACTCTGCTCTCTGCGCAATTC 708
DB 181 TCCTCTCTCTCTCTGCGCACCGCCACCAACCGCACTCTGCTCTCTGCGCAATTC 240

QY 709 ATTGATGTGTCGGGAGGCGGACGCTCCCATTCGCCATTTCTACCGAGTATTGAGAGT 768
DB 241 ATTGATGTGTCGGGAGGCGGACGCTCCCATTCGCCATTTCTACCGAGTATTGAGAGT 300

QY 769 CAACACATTTCTTAAACCGATGGAATGAGAGACATTCGAAAGCTAGGGTTTGGAAACG 828
DB 301 CAACACATTTCTTAAACCGATGGAATGAGAGACATTCGAAAGCTAGGGTTTGGAAACG 360

QY 829 CCGCAATTCGGTTTCAAGGACGACGCTTTAATCAGCCGAGACGATTTCTCAAGCTGCT 888
DB 361 CCGCAATTCGGTTTCAAGGACGACGCTTTAATCAGCCGAGACGATTTCTCAAGCTGCT 420

QY 889 TGGGAACTCTGTATATCTCGGCTTAAAGAGAGATGATGAAAGGCTTCTTGATGAT 948
DB 421 TGGGAACTCTGTATATCTCGGCTTAAAGAGAGATGATGAAAGGCTTCTTGATGAT 480

QY 949 GAAGAAGCTACAGTCACTGATGTTCTTGGGATTAAGTAATTCGATTCGGAATTA 1008
DB 481 GAAGAAGCTACAGTCACTGATGTTCTTGGGATTAAGTAATTCGATTCGGAATTA 516

QY 1009 TAAAGTTTCTGTTTAAATTTCAATGAAATGAAAGAACTTTATCTAGTGA 1068
DB 517 TAAAGTTTCTGTTTAAATTTCAATGAAATGAAAGAACTTTATCTAGTGA 1128

QY 1069 GGTTCCTGGGGCTCTGTGTATGCAAGAAAGTGTGAGACTGAGATGATTTCTGGGT 1128
DB 519 GGTTCCTGGGGCTCTGTGTATGCAAGAAAGTGTGAGACTGAGATGATTTCTGGGT 578

QY 1129 TGGTGAAGCTCTGCTTAAGAGAGGTGCTTAAGTCTTTAAGCAAGATGTGTTAGT 1188
DB 579 TGGTGAAGCTCTGCTTAAGAGAGGTGCTTAAGTCTTTAAGCAAGATGTGTTAGT 638

QY 1189 TATGGCGCTTGGCTTCTGATGTCTGAGGGATGCTATGGCATTTGATTCACCTGATTT 1248
DB 639 TATGGCGCTTGGCTTCTGATGTCTGAGGGATGCTATGGCATTTGATTCACCTGATTT 698

QY 1249 TATTACTGTTTATGAGTTTGTGAGAAAGTGTGAAGCTTTTAAAGATGTTGACTTGC 1308
DB 699 TATTACTGTTTATGAGTTTGTGAGAAAGTGTGAAGCTTTTAAAGATGTTGACTTGC 742

QY 1309 TTTGTAATTTGACGAGCGTTGCTTATTAAGAACTTCTTATGATTGATTTGTTAT 1368
DB 743 TTTGTAATTTGACGAGCGTTGCTTATTAAGAACTTCTTATGATTGATTTGTTAT 742

QY 1369 GAGCTTGTGTAGAGAGAGCAAGTACCTTGACCGGATTTACGTGCAAAATTTGA 1428
DB 1369 GAGCTTGTGTAGAGAGAGCAAGTACCTTGACCGGATTTACGTGCAAAATTTGA 1428

DB 743 -----AGAGAGAGAGAGCAAGTACCTTGACCGGATTTACGTGCAAAATTTGA 791
QY 1429 TGAGACTTTGGAAGAGATCACTCGCGTTATGTCTTGAAGCTATTGGCTTACCGCTTG 1488
DB 792 TGAGACTTTGGAAGAGATCACTCGCGTTATGTCTTGAAGCTATTGGCTTACCGCTTG 851

QY 1489 TGATGATTACGCTCGGAAAAAGACTAAATGGTTTAAAGCGGTGCGGAATATTTTGCTC 1548
DB 852 TGATGATTACGCTCGGAAAAAGACTAAATGGTTTAAAGCGGTGCGGAATATTTTGCTC 911

QY 1549 TGTGGAG 1556
DB 912 TGTGGAG 919

RESULT 2

LOCUS

DEFINITION

CNS09255 741 bp mRNA linear HTC 06-FEB-2004
Arabidopsis thaliana Full-length cDNA Complete sequence from clone
GSLTILJ12E09 of Silique of strain col-0 of Arabidopsis thaliana
(thale cress).

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

Submitted (18-NOV-2003) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrel@genoscope.cns.fr
Genoscope.
Direct Submission
Unpublished
2 (bases 1 to 741)

JOURNAL

AUTHORS

COMMENT

The sequences are based on single pass reads.
Life Technologies (a division of Invitrogen) members carried out
full-length libraries construction : Temple G.
Genoscope members carried out sequencing and annotation : Castellii
V., Aury J.M., Jallion O., Wincker P., Menard M., Cruaud C.,
Schachter V., Weissenbach J., Salanoubat M.
URGV INRA : Clepet C., Caboche M.
Annotation is based on the June 2003 version of the Arabidopsis
genome released by MIPS (Munich Information center for Protein
Sequences). 5 prime and 3 prime are assembled with Phrap.
http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/Full
length
http://www.genoscope.cns.fr/cgi-bin/ggb/ggb?source=Arabidopsis.
Location/Qualifiers

FEATURES

source

1. 741
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Col-0"
/db_xref="taxon:3702"
/clone="GSLTILJ12E09"
/tissue_type="Silique"
/plasmid="pCMVSPORT_6"
1. 741
/gene="At5g42480"

ORIGIN

Query Match 11.9%; Score 436.2; DB 3; Length 741;
Best Local Similarity 83.4%; Pred. No. 3.4e-101;
Matches 548; Conservative 0; Mismatches 18; Indels 91; Gaps 1;

QY	2970	TTTCTTTCCTTGATATTTTTCCTTTTCATTTAGAGGTGAGTCAGAGCTACAGATTCAGAA	3029
Db	176	TTTCTTCTATGGAATCTGATGTCGTACCAATAGGGTCAGTCAGACTACGAATTCAGAG	235
QY	3030	CACCTCCAGATGAGATGCTAGAGCTGACAGAAATATAGTATCCAGAGTCAGAGATTA	3089
Db	236	CACCTCCAGATGAGATGCTAGAGCTGACAGAAATATAGTATCCAGAGTCAGAGATTA	295
QY	3090	AGTCTCTGGCTTTTGGGCTGTGATCAACGCAATAGAAATTTAACAGAGGTGAGGAAATAA	3149
Db	296	AGTCTCTGGCTTTTGGGCTGTGATCAACGCAATAGAAATTTAACAGAGGTGAGGAAATTA	340
QY	3150	TCTACAAATCAATCAATATGTGTGAAACATGTGACATGATTAATAGTGTGGCTGTGT	3209
Db	341	-----	340
QY	3210	TGATTTCTGTTATTTATAGTTAGTTGGATGGGCGAAATGCTGAAGATTTTGGACTGACAGCA	3269
Db	341	-----AGTTTGGATGGGCGAAATGCTGAAGATTTTGGACTGACAGCA	384
QY	3270	GCTGAAACCTGGCGAGCTTGGGTTGGTTATGATTAATACATGTTGAAACATATCTGTGGAC	3329
Db	385	GCTGAAACCTGGCGAGCTTGGGTTGGTTATGATTAATACATGTTGAAACATATCTGTGGAC	444
QY	3330	AGTGTGACAGTCTCAGAGATGGAACCCGCTCTGGTGAAGAACATCTGAGAGAGCT	3389
Db	445	AGTGTGACAGTCTCAGAGATGGAACCCGCTCTGGTGAAGAACATCTGAGAGAGCT	504
QY	3390	GCTGTCTATCTGATTTGGTTCATCCAGAAAACATGCTAGTGTGACAACTTACACA	3449
Db	505	GCTGTCTATCTGATTTGGTTCATCCAGAAAACATGCTAGTGTGACAACTTACACA	564
QY	3450	ACAAGATACGAAGTTTCTGGTCCAGTCAAGGTGGAATACTGAAAGCTCTGTCTT	3509
Db	565	ACAAGATACGAAGTTTCTGGTCCAGTCAAGGTGGAATACTGAAAGCTCTGTCTT	624
QY	3510	GCATCATTAATTAATCTCATATGATGACATGTCTGAGCTTGGAGATTTCTTTGTTCTGTA	3569
Db	625	GCATCATTAATTAATCTCATATGATGACATGTCTGAGCTTGGAGATTTCTTTGTTGTA	684
QY	3570	ATTCTCTCTCTAGTGTAGTGTATTAATGAACAACAAAAATTAAGCTTGGCA	3626
Db	685	ATTCTCTCTCTAGTGTAGTGTATTAATGAACAACAAAAATTAAGCTTGGCA	741
RESULT 3			
CNS09YNM		741 bp	linear
LOCUS			HTC 06-FEB-2004
DEFINITION		Arabidopsis thaliana Full-length cDNA complete sequence from clone	
		GS1TSL532A05 of Siliques of strain col-0 of Arabidopsis thaliana	
		(thale cress).	
ACCESSION		BX833489	
VERSION		BX833489.1	GI:42455179
KEYWORDS		HTC; GSUT cDNA.	
SOURCE		Arabidopsis thaliana (thale cress)	
ORGANISM		Arabidopsis thaliana	
		Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
		Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;	
		rosids; euroside II; Brassicales; Brassicaceae; Arabidopsis.	
REFERENCE		1 (bases 1 to 741)	
AUTHORS		Castelli,V., Aury,J.M., Jallion,O., Wincker,P., Clepet,C.,	
		Mendel,M., Cruaud,C., Quetier,F., Scarpelli,C., Schachter,V.,	
		Temple,G., Caboche,M., Weissenbach,J. and Salanoubat,M.	
		Whole Genome Sequence Comparisons and 'Full-length' cDNA Sequences:	
		A Combined Approach to Evaluate and Improve Arabidopsis Genome	
		Annotation	
JOURNAL		Unpublished	
REFERENCE		2 (bases 1 to 741)	
AUTHORS		Genoscope.	
TITLE		Direct Submission	
JOURNAL		Submitted (18-NOV-2003) Genoscope - Centre National de Sequenage :	
		BP 131,91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr	
		- Web : www.genoscope.cns.fr)	

COMMENT	ORIGIN	FEATURES	source
The sequences are based on single pass reads. Life Technologies (a division of Invitrogen) members carried out full-length libraries construction : Temple G. Genome members carried out sequencing and annotation : Castelli V., Aury J.M., Jailion O., Wincker P., Menard M., Cruaud C., Schachter V., Weissenbach J., Salanoubat M. UNIV INRA : Clapet C., Caboche M. Annotation is based on the June 2003 version of the Arabidopsis genome released by MIPS (Munich Information center for Protein Sequences). 5 prime and 3 prime are assembled with Phrap. http://www.genoscope.cns.fr/externe/sequences/Banque_Projet_EF/FullLength	Query Match 11.9%; Score 434.6; DB 3; Length 741; Best Local Similarity 83.3%; Pred. No. 8,7e-101; Matches 547; Conservative 0; Mismatches 19; Indels 91; Gaps 1;	Location/Qualifiers 1..741 /organism="Arabidopsis thaliana" /mol_type="mRNA" /strain="Col-0" /db_xref="taxon:3702" /clone="GSIRESIL592A05" /issue_type="Siliques" /plasmid="pCMVSPORT_6" 1..741 /gene="At5g42480"	gene
3570 ATTCTCTCTAGTAAAGTTTATTAAGACAACAAAAATTAAAGTCTTGGTA 3569	3570 TTTCTTCCCTGATTTTTTTCTTTTGCAATTTAGGCTCAGAGCTGACGATTCAGAAG 3029		
685 ATTCTCTCTAGTAAAGTTTATTAAGACAACAAAAATTAAAGTCTTGGTA 741	176 TTTCTTCTATGGAATCGATGTGCTACCAATAGGCTAGTCAGAGCTGACGATTCAGAAG 235		
	3030 CACTTCCAGAAATGATGCTAGAGCTCAGAGCATATAGTATCCAGTGGCAGAAATTA 3089		
	236 CACTTCCAGAAATGATGCTAGAGCTCAGAGCATATAGTATCCAGTGGCAGAAATTA 295		
	3090 AGTCTCTGCTTTTGGGCTGATCAACCGCATAGAAATGTTACAGAGGTGAGGAATAA 3149		
	296 AGTCTCTGCTTTTGGGCTGATCAACCGCATAGAAATGTTACAGG----- 340		
	3150 TCTACAAATCAATCAATGTTGTTGAAAAGTGTGACATGATTTATGTCGTGCTTGT 3209		
	341 ----- 340		
	3210 TGATTCGTATTATATAGTTTGGATGGGCGAATGCTGAAGATTTGGACTGACAGACA 3269		
	341 -----AGTTTGGAGGGCGAATGCTGAAGATTTGGACTGACAGACA 384		
	3270 GCTGAAGATCGAGAGCTTGGGTTGTTATGATTATCACTGTTGAAATCTGTTGAC 3329		
	385 GCTGAAGATCGAGAGCTTGGGTTGTTATGATTATCACTGTTGAAATCTGTTGAC 444		
	3330 AGTGTGACAGTCTCAGCAGATGAAACCCGCTCTGTTGAGCAACTCTGAGAGACT 3389		
	445 AGTGTGACAGTCTCAGCAGATGAAACCCGCTCTGTTGAGCAACTCTGAGAGACT 504		
	3390 GCTTGTCTATCTGATTTGGTTATCCAGAAACAAAGCTATCTGATGCAACCTTACA 3449		
	505 GCTTGTCTATCTGATTTGGTTATCCAGAAACAAAGCTATCTGATGCAACCTTACA 564		
	3450 ACAAGATACGAAGTTTCTGGTCCAAGTCAAGGTTGCAAAATCACTGAAGGCTCTGTTCT 3509		
	565 ACAAGATACGAAGTTTCTGGTCCAAGTCAAGGTTGCAAAATCACTGAAGGCTCTGTTCT 624		
	3510 GCATCAATATATATCAATATGATGATGTCAGCTTGCAGATTCTCTTGTGTTGTA 3569		
	625 GCATCAATATATATCAATATGATGATGTCAGCTTGCAGACTCTCTTGTGTTGTA 684		

	RESULT 4	A1998415/c	561 bp	mRNA	linear	EST 08-SEP-1999
LOCUS	A1998415/c					
DEFINITION	701545606 A. thaliana, Columbia Col-0, rosette-2 Arabidopsis thaliana cDNA clone 701545606, mRNA sequence.					
ACCESSION	A1998415					
VERSION	A1998415.1					
KEYWORDS	EST.					
SOURCE	Arabidopsis thaliana (thale cress)					
ORGANISM	Arabidopsis thaliana					
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids; euroside II; Brassicales; Brassicaceae; Arabidopsi.					
AUTHORS	1 (bases 1 to 561) Chen,Y., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D., Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brozka,P., Gorgone,G., Burns,D., Griffin,J., Mouanoutou,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ion,B., Kasebury,K., Borillo,C., Caprio,T., Policky,J., Suzuki,G., Asentine,C., Shah,S., Nobrifa,A., Murry,L., Turner,C., Kirkorian,S., Elder,L. and Hanson,D.					
TITLE	Arabidopsis thaliana Gene Expression Microarray					
JOURNAL	Unpublished (1999)					
COMMENT	Contact: David Smoller, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 4633 World Parkway Circle, St. Louis, MO 63134, USA Tel.: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com. Location/Qualifiers 1..561 /organism="Arabidopsis thaliana" /mol type="mRNA" /ecotype="Col-0" /db_xref="taxon:3702" /cloned="701545606" /tissue_type="rosette" /dev stage="4 - 7 weeks" /clone_1ib="A. thaliana, Columbia Col-0, rosette-2" /note=Vector: pSPORT; Site 1: NotI; site 2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. CDNA synthesis was initiated using a Notti-oligo(dT) primer. Double-stranded cDNA was bluntend, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and Sal sites of the pSPORT vector."					
ORIGIN						
Query Match	11.0%; Score 402.2; DB 1;					
Best Local Similarity	82.2%; Pred. No. 1.9e-92;					
Matches 514; Conservative	0; Mismatches 20; Indels 91; Gaps 14					
Dn	2970 TTTCTTCCTTGATTTTTTTCTTCGATTAGGGTCAGTCAGACTGCAGATTCAAGA	3028				
Dn	534 TTCTTCCTCANTGAANNCTAGTCGTCCACCATCGGTGAGTCAGACTGACATTCAAG	475				
Dn	3030 CAC TTC CCAG ATG GA CTG AGA CTC GAG CA GA AA TA TGTA TC CA NG TG GC GA GA GT TA	3089				
Dn	474 CAC TTC CGA NTG CAT GCT TAG AGA CTC GAG CA GA AA TA TGTA TC CA NG TG GC GA GA GT TA	415				
Dn	3090 AGT CTC GG CTT TTT GGG CCT GAT CA TCG CGC ATA GAA ATG TTA CCA GAG TGA GGA AT AAA	3145				
Dn	414 AGT CTC GG CTT TTT GGG CCT GAT CA TCG CGC ATA GAA ATG TTA CCA G-	370				
Dn	3150 TCTACA ATTCAATCAATGTGTGTGAATAAAGCTTTGACAGATTAATAGTGTGGTCTTT	3205				
Dn	369 -----	370				

QY	3210	TGATCTGTTATTTATTAAGTTTGGAAATGGCGAAATGCTGAAGATTTGGACTGCACAGACA	3269
Db	369	-----AGTTTTGGATGGGGCAATGCTGAGATTTTGGACTGCACAGACA	326
QY	3270	GCTGAAATCTGCGACAGCTTGGGTGGTTTATGATTATACATGTTGAAACTATCTGTTGAC	3329
Db	325	GCTGAAATCTGCGACAGCTTGGGTGGTTTATGATTATACATGTTGAAACTATCTGTTGAC	266
QY	3330	AGTGTGACAGTCTTCAGACAGATGAAACCCGTCCTTGGTGGAAAGCACTCTGGAGAGATCT	3389
Db	265	AGTGTGACAGTCTTCAGACAGATGAAACCCGTCCTTGGTGGAAAGCACTCTGGAGAGATCT	206
QY	3390	GCTTGTCTATCTGATTTGGTTCATCCAGAAACAAATGCTACTGATGTCAAGAACTTACACA	3449
Db	205	GCTTGTCTATCTGATTTGGTTCATCCAGAAACAAATGCTACTGATGTCAAGAACTTACACA	146
QY	3450	ACAAGATACGAAGTTTCTGGTCCAAATCTAGGGTGGAAAAATCACTGAAAGGCTCTGTTCTT	3509
Db	145	ACAAGATACGAAGTTTCTGGTCCAAATCTAGGGTGGAAAAATCACTGAAAGGCTCTGTTCTT	86
QY	3510	GCATCATTAATATACCTCATATGTAGACATGTCTGAGCTTGCAGATTCCTTTGTTCTGTAA	3569
Db	85	GCATCATTAATATACCTCATATGTAGACATGTCTGAGCTTGCAGATTCCTTTGTTGTAA	26
QY	3570	ATTCTCTCTCTAAGTTAGTGTATT	3594
Db	25	ATTCTCTCTCTAAGTTAGTGTATT	1

RESULT 5					
B2437564/c					
LOCUS	B2437564	785 bp	DNA	linear	GSS 13-DEC-2002
DEFINITION	BONNRN72TR BO_1.6_2_KB_tot Brassica oleracea genomic clone BONNR72,				

ACCESSION	BZ437564	GI:26691135
VERSION	BZ437564.1	
KEYWORDS	GSS.	
SOURCE	Brassica oleracea	
ORGANISM	Brassica oleracea	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 785)	Town,C.D., Van Aken,S., Uteirback,T., Koo,H. and Fraser,C.M.	Whole genome shotgun sequencing of Brassica oleracea	Unpublished (2001)	Other_GSSs: BONRN72TF

9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtowm@tigr.org
DNA is from a doubled haploid provided by Tom Osborn
Seq primer: TR
Class: sheared ends.

FEATURES	source	Location/Qualifiers
1..785		
/organism="Brassica oleracea"		
/mol_type="genomic DNA"		
/strain="NO1000DH3"		
/db_xref="taxon:3712"		
/clone="BONEN72"		
/clone_1lb="BO_1.6_2_KB tot"		
/note="vector: pHS01; Site_1: total DNA inserted into pHS01 using BstXI linkers"		
ORIGIN		

Query Match	9.9%	Score 363.6;	DB 8;	Length 785;
Best Local Similarity	73.3%	Pred. No. 2.1e-82;		
Matches 582; Conservative	0;	Mismatches 109;	Indels 103;	Gaps 5;


```

Qy 1061 CTAGTGAAGTCTCGGGGCTCTCTGTATATGCAAGAGGTGTGAGTGAATAGT 1120
Db 699 CAATGAAGTCTCGGTGCTCTGTGTACTGCAAGAGTGTGAGTGAATAGT 640
Qy 1121 CTTGGGGTGTGAGGCTCTGCTTAAGAGAGTTCCTTAAGTCTTAAGCAAGATGTG 1180
Db 639 CTTGGTGAAGAGAGCTTGTCTTAAGAGAGTGTGCTTAAGTCTTAAGCAAGATGTG 580
Qy 1181 GTTTAGTGAAGGCTGCTGCTTCTGATGCTGAGAGAGTGTGATGAGTGAATGCA 1240
Db 579 GTTTGTGAAGGCTGCTGCTTGTGAATCTTCAGAGAGTGTGATGAGTGAATGCT 520
Qy 1241 CCTGATTTATTAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1300
Db 519 CTTGATTTATTAAGTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 471
Qy 1301 TGACTTGTGTTGTAATTTGACGAGCGTGTATTAAGAACTTTTGTGATTTGAT 1360
Db 470 ----- 471
Qy 1361 TTGTATTTAGTCTTGTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1420
Db 470 -----TGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
Qy 1421 CAATTTGATGAGCTTGTGAAGAGATCACTCGGCTGTATGCTTGTGAGTGAATGCT 1479
Db 426 CAGTTGATGAGCTTGTGAAGAGATCACTCGGCTGTATGCTTGTGAGTGAATGCT 367
Qy 1480 ACCGCTTGTGATGATTAAGCTGCGAAAAGATTAATGTTTGAAGCGTGTGCGAATAT 1539
Db 366 ACCGCTGCGGAGATGAT-----AAACGACAGATGTTTGAAGCGTGTGACGAAATAT 316
Qy 1540 TTTGTGTCTGTGTGAAGAGAGTGTGAGTCACTCTTGTGTGGGGTGTGACCGTGA 1599
Db 315 CTTGTGTCTGTGTGAAGAGAGTGTGAGTCACTCTTGTGTGGGGTGTGACCGTGA 256
Qy 1600 GTTATGATGAGAGCGTTTGAAGATGAGAGTGTGAGAGAGTGTGAGAGTGTGAGATAC 1659
Db 255 ATTTATGATGAGAGCGTTTGAAGATGAGAGTGTGAGAGTGTGAGAGTGTGAGATAC 196
Qy 1660 CTTTGTATTTCTTGAAGATGATTAATTAAGTGTGATTTTGAAGTGTGATTTTGA 1714
Db 195 TTTTGTATTTCTTGAAGATGATTTTGAAGTGTGATTTTGAAGTGTGATTTTGA 136
Qy 1715 TTTGTGTGTGATGATTTTGTGATGATGATGATGATGATGATGATGATGATGATG 1774
Db 135 TTTGTGTGTGATGATTTTGTGATGATGATGATGATGATGATGATGATGATGATG 76
Qy 1775 AAGTTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1833
Db 75 AAGTTTGAAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 16
Qy 1834 CTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1847
Db 15 CTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2

```

RESULT 6
 BP620404/c 434 bp mRNA linear EST 26-JUN-2004
 LOCUS BP620404 RAPL16 Arabidopsis thaliana cDNA clone RAPL16-37-C18 3',
 DEFINITION mRNA sequence.
 ACCESSION BP620404
 VERSION BP620404.1 GI:49271586
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (chale crese)
 ORGANISM Arabidopsis thaliana
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 434)
 REFERENCE Seki M., Narusaka M., Kamiya A., Ishida Y., Satou M., Sakurai T.,
 Nakajima M., Enju A., Akiyama K., Oono Y., Muramatsu M.,

```

TITLE Hayashizaki Y., Kawai J., Carninci P., Itoh M., Ishii Y.,  

JOURNAL Arakawa T., Shibata K., Shinagawa A. and Shinzaki K.  

MEDLINE Functional annotation of a full-length Arabidopsis cDNA collection  

PUBMED Science 296 (5565), 141-145 (2002)  

COMMENT 21932900  

11910074  

Contact: Motoaki Seki  

Plant Functional Genomics Research Group  

RIKEN Genomic Sciences Center  

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  

Tel: 81-298-36-4359  

Fax: 81-298-36-9060  

Email: msek@rtc.riken.go.jp  

reversed clone; please visit our web site  

(http://efgweb.gsc.riken.go.jp/) for further details.

```

FEATURES
 source
 1..434
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAPL16-37-C18"
 /lab_host="DH10B"
 /clone_id="RAPL16"
 /note="Site_1: BamHI; Site_2: SalI; dark-grown"

ORIGIN
 Query Match 9.8%; Score 359.4; DB 5; Length 434;
 Beest Local Similarity 98.0%; Pred. No. 2.2e-81;
 Matches 385; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

```

Qy 3224 ATAGTTTGTGATGGCGAATGCTGAAGATTTGAGTGAAGAGAGAGAGAGAGAGAGAG 3283
Db 402 AGAGTTTGTGATGGCGAATGCTGAAGATTTGAGTGAAGAGAGAGAGAGAGAGAG 343
Qy 3284 GCTTGGGTGTTGATGATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 3343
Db 342 GCTTGGGTGTTGATGATTAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 283
Qy 3344 AGCAGATGGAACCCGCTGCTGTGTGAAGCACTGTGAAGAGTGTGTTCTATCTGA 3403
Db 282 AGCAGATGGAACCCGCTGCTGTGTGAAGCACTGTGAAGAGTGTGTTCTATCTGA 223
Qy 3404 TTTGTTTCATCCAGAAAACAATGCTATGATGTGAGAACTTGAACAACAATGAGAGT 3463
Db 222 TTTGTTTCATCCAGAAAACAATGCTATGATGTGAGAACTTGAACAACAATGAGAGT 164
Qy 3464 TTTGTGTCAAGTCAAGGTGGAATATCA-CTGAAGGCTGTTCTTGATCATATATTA 3522
Db 163 TTTGTGTCAAGTCAAGGTGGAATATCAAGTGAAGGCTGTTCTTGATCATATATTA 104
Qy 3523 CTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3582
Db 103 CTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44
Qy 3583 GTTAGTGTATTAATGAACAAAATAATTA 3615
Db 43 GTTAGTGTATTAATGAACAAAATAATTA 11

```

RESULT 7
 CD573714 897 bp mRNA linear EST 12-JUN-2003
 LOCUS CD573714
 DEFINITION UCRP101_01_F12_T3 Poncirus trifoliata CTV-challenged cDNA library -
 UCR Poncirus trifoliata cDNA clone UCRP101_01_F12, mRNA sequence.
 ACCESSION CD573714
 VERSION CD573714.1 GI:31669616
 KEYWORDS EST.
 SOURCE Poncirus trifoliata
 ORGANISM Poncirus trifoliata
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
 1 (bases 1 to 897)

[illegible]

DEFINITION	MBEKG28T8B_mch2_Medicago truncatula genomic clone 73F8, genomic survey sequence.
ACCESSION	CG961431
VERSION	CG961431.1
KEYWORDS	GI:39883077
SOURCE	GS.
ORGANISM	Medicago truncatula (barrel medic)
REFERENCE	Medicago truncatula
AUTHORS	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosoid 1; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
JOURNAL	1 (bases 1 to 954)
COMMENT	Town, C.D., Sherry, J., Koo, H. and Feldblyum, T.F. Sequencing of BAC ends from Medicago truncatula Unpublished (2003)
	Other_GSSs: MBEKG28T8B
	Contact: Chris Town
	TIGR
	9712 Medical Center Drive, Rockville, MD 20850, USA.
	Tel: 301-838-3523
	Fax: 301-838-0208
	Email: cdtown@tigr.org
	Seq primer: CAGGAACACGCTATGACC
	Class: BAC ends.
FEATURES	Location/Qualifiers
source	1..954
	/organism="Medicago truncatula"
	/mol_type="genomic DNA"
	/culivar="genotype A17"
	/db_xref="taxon:3880"
	/clone="73F8"
	/clone_lib="mth2"
	/note="Vector: pGelBAC11; Site_1: HindIII; Site_2: HindIII; Cook, D.R. and Kim, D.J., unpublished"
ORIGIN	
Query Match	8.3%; Score 305.6; DB: 9; Length 954;
Best Local Similarity	65.8%; Pred. No. 2.3e-67;
Matches 563; Conservative	0; Mismatches 239; Indels 54; Gaps 6
Dn	1377 TGTAGGAGGAAGAGCAAGTAGCCCTTCACCGGATTTACGTGCAAAATTGATGACACTT 1436
Dn	856 TGTAGGAAGAAAGGGGCAAGACGCTTAGCACCGGATTTACAAACAAATTGATGACAC 797
Dn	1437 TGTAGGAGATCATCTCCGCGTTATGCTTGGAGCTACTTGGCTTACCGCTTGGTATGATT 1496
Dn	796 TTGAAGAGATPAACCTCACGTTGTATTTAGAACTTTAGCCCTTCCTTATGATGAC 737
Dn	1497 ACGGTGGAAGAAAGCTAAATGATTAAAGCGGTGCGGAATATTTTGGTCTGTGGAG 1556
Dn	736 ATCAGAGCGGAGAGGAAGAAAGGTCTCCAGGTGTCCGCAATTTCTGGGCAATTGGAG 677
Dn	1557 GAGGTGAGCATCAGCTCTTGTGGGGGATTTGACCCGCGAAGTTTATGATGAGCGGT 1616
Dn	676 GTGAGGAGGACGACGAAATGCTGGAGTTTCAACGAGGAGATTATGATGATGAGCAT 617
Dn	1617 TTTTACGAATGACAGCTGCTGAGCAGGATTAACAGTTAGATA-----CCTTTTTTAT 1670
Dn	616 TCCTGCATATGAAAGCTGCGGAACAGGTGTTATAGAAATATATGTCATGTTTTTAT 557
Dn	1671 TTCTTTGAGATGATATATCTTTAGTTTCTCATTTATATGAT----- 1713
Dn	556 TTAAATTTTCTCGCCCCCTTAATTTTTTGTATTAAGAGTTTATACATTTATCTCTTCA 497
Dn	1714 -----GTTGTGGTAGTTGATCTTTTGTAGTCAACCCAG 1751
Dn	496 TTAAATCTTTTATCTCAATCTTTTTTTTTTAAAGTGAACCTTTTTTGAAGCAACCTAG 437
Dn	1752 CAATATTCGACGAGATCATTTGAAGTTTACGAAGTTGACCTTCTTGGCTCAAGC 1811
Dn	436 CAATATTCGACGAGATTTTGAAGGCTTAAGGGGTTGACCTTCATTTGATGTTGCAAGC 377
Dn	1812 TTTTATTTGTAAGAGCACACCTTTTACAGAGATGCTGATAGCAATTCAGCAACTTCA 1871

Db	376	CTTTGTAGTAAAAAGCCACATCTTATCCAGATGCTGATAATTATTATTCATCAACTTCA	317
Qy	1872	GCAGGCTTAAGGTAAATGCGCTATGAGAAVTCCTGCAGTGTGTATGATAC-----ACGGAA	1925
Db	316	ACAAACTAAGGTAAACAATATGAGGAATGTCCTCCCTGTATTATACCTCCATGAGAGATGA	257
Qy	1926	TAATTGGAGATAGACTTCGCGTCTAGAAAAGGGGACTCTGTGCACCTGCTTATAGCAAACT	1985
Db	256	GAAAGAGAAATTTGATTTTGATTAAGAAAGGGGCTGTGTGCACCTGCTGTGTGGGGAGCT	197
Qy	1986	TGATGAATGCCGTATGCTGTGGGCTTAGACAGTGA-GGATTCACAATATAGGAATCCAG	2044
Db	196	TGATCAATGTGATCATTGTTGGGCTCTCGATGATGACAGTCTCACCTTATGAAAACCAT	137
Qy	2045	CTATTGGAGATTTGTTTGGAGAAATCAAAATCGGATG-ACAAATGATATCTCCCTGGA	2103
Db	136	CTATTATATGACTTATATATATGAAAAACGAAAAAGTGATGAAGATACAGTATCTTCTGGA	77
Qy	2104	CTATGCAAAATGTTGGAAACCTGGTTGGCAGGGGTTGCTTTC-CTAGGTTCAAGACAC	2162
Db	76	CTCTGTAAATTTGTGGAGACATGGTTGATGAGAGGTGTTTCTCTTAGGTTTAGAGATAC	17
Qy	2163	CAAGATATAAAAATTT 2178	
Db	16	TAAAGACACAACTTT 1	

RESULT 12	LOCUS	DEFINITION
BU046755	631 bp	mRNA linear EST 26-AUG-2002
PP_LBA0027104f		Peach developing fruit mesocarp Prunus persica CDNA
clone_PP_LBA0027104f		mRNA sequence.

ACCESSION	BU046755
VERSION	BU046755.1
KEYWORDS	GI:22486832
SOURCE	EST.
ORGANISM	<i>Prunus persica</i> (peach)
	<i>Prunus persica</i>

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 631)	Callaghan, A., Palmer, M., Main, D., Wing, R. and Abbott, A.	Peach Model Genome for Rosaceae	Unpublished (2002)	Contact: Abbott, A. Part of <i>Genetics and Biochemistry</i>

FEATURES	Location/Qualifiers
SOURCE	1. 631

```

/clone.lib="Peach developing fruit mesocarp"
/notes=vector: pbluescript II SK(-); Site 1: EcoRI;
Site 2: XhoI; authority=Prunus persica L. Batsch, the
sequence has been trimmed to remove vector sequence and
contains a minimum of 100 bases of phred value 20 or
above. For more details on library preparation and
sequence analysis go to
http://www.genome.clemson.edu/projects/peach. To order
this clone go to http://www.genome.clemson.edu/orders

```

ORIGIN

Query March	8.1%;	Score 295.6;	DB 5;	Length 631;
Best Local Similarity	71.7%;	Pred. No. 7.8e-65;		
Matches 407; Conservative	0;	Mismatches 149;	Indels 12;	Gaps 1;

QY	1708	ATGATATGTGTGAGTGTGATCTTTTGTGACATCCCAAGCAATATTCAGCAG	1767
Db	64	ATGATCTCAGCTGACAGGTGATTATTTGTAGCTACCCCAAGTATATCCCGCAGAA	123
QY	1768	TCAATTTGAAGTTTACGAAGTTGCACTTGCTCTGTGTGCTCAAGCTTTATTTGTAAG	1827
Db	124	AGCTTTGAAGTTTATGAGGGTGGCTCTTGCCGCTGTGCTCAAGCTTTGTGTGAAAAA	183
QY	1828	CCACACCTTTTACAGAGATGCTGATTAACCAATTCAGCACTTCAGCAGCGTAAGSTATG	1887
Db	184	CCTATATCAATTCAGAAATGCTGAAAACCTTATCCAGAAACTTCAGCAGCTTAAGGTACA	243
QY	1888	GCTATGAGAGATTCCTGCGATGTTGTATGATATCACGAATATATGGAGATATGACTCGGT	1947
Db	244	GCTGTAGACATTTCTTTGACACATATATACCAAGAAAGCAGTGAATGACTTTGCT	303
QY	1948	CTAGAAAGGGGACTCTGTGCACTGCTTATATAGGCAAGTTGATGAAATGCCGTATGTGTTG	2007
Db	304	TTGGAAGAGGGGACTCTGTACACTTTCTTAGAGGAGCCTTGATGACAGTCGTTCTGTGTTG	363
QY	2008	GGCTTTAGCAGTGAAGATTCACATATATAGGAATCAAGCTATTTGTGAGTTGTTTGGAG	2067
Db	364	GGCTTAAACAGTATATGATTCACCATATATGAAGATCACTGTGTTGAGCTTTGTCTTGGAG	423
QY	2068	AATTCAAATGCTGATG-----ACAATGATATCCCTGAGACTATGCAAAATG	2115
Db	424	AATCTAAAGAGATGACATGACATATGACATATATCTTCTGAGCTTTGCAAGCTA	483
QY	2116	TTGGAAACCTGTGTGGCAGGGGTGTCTTTCTTAGTTGACAGACACCAAGATTAATAA	2175
Db	484	TTGGAGAGGTGTGTGATGAGAGGTGTATTCCTCCAGGTTTGAAGACACCAAGACATAGAG	543
QY	2176	TTTAAACTCGGGGAGCTATGATGATATCTATGTGTTTGAATTACTTGGAAAGAGTGGAG	2235
Db	544	TTTCAGACTGGGAGACTATGATGATATCTTCAAGTCTTGAGATACTTAAAGAAAGCCTGAT	603
QY	2236	GTAGTTAGGGTCTCCTTAGAGTGGC	2263
Db	604	GGCACTATATGTTCAACCTTAGTGTGCTG	631

RESULT	13
LOCUS	CO079829
DEFINITION	CO079829 832 bp mRNA linear EST 15-JUN-2004 GR_Ea42007.r GR_Ea Gossypium raimondii cDNA clone GR_Ea42007 3 , mRNA sequence.

ACCESSION	CO079829	
VERSION	CO079829.1	GI:48749310
KEYWORDS	EST.	
SOURCE	<i>Gossypium raimondii</i>	
ORGANISM	<i>Gossypium raimondii</i>	

REFERENCE
AUTHORS
KIM, H., YU, Y., KUDRINA, D., HATFIELD, J., STUM, D., MUELLER, C.,
UDALL, J. A., KAPP, R. A., WENDEL, J. F., RAO, K., SODERLUND, C. and
WING, R. A.
TITLE
Global assembly of Cotton ESTs
JOURNAL
Unpublished (2004)
COMMENT
Contact: Rod A. Wing;
Direct: Rod.A.Wing@ars.ars-gov

Arizona Genomice Institute
The University of Arizona
Pohres Building Room 303, Tucson, AZ, 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: <http://genome.arizona.edu>

Plate: 42 row: 0 column: 07.

FEATURES
Location/Qualifiers
1. .832

/organism="Geosyrium raiondii"
/mol_type="mRNA"
/db_xref="taxon:29730"
/clone="GR_Ba42007"
/issue_type="whole seedlings"
/dev_stage="first true leaves"
/lab_host="DH10B"
/clone_lib="GR_Ba"
/note="vector: pCMV.SPORT-6.1; Site_1: NotI; Site_2: EcoRV; Library made by Invitrogen with RNA supplied by Wendle Lab. Directional cloned into NotI-EV. Colonies plated/picked by AGI. More glycerol clones held in -80."

ORIGIN

Query Match 8.0%; Score 293.4; DB 7; Length 832;

Best Local Similarity 69.2%; Pred. No. 3.2e-64;

Matches 424; Conservative 0; Mismatches 171; Indels 18; Gaps 1;

QY 1757 TTCACGACGAGTCATTGAAGTTACGAAGTTCGACTTCTTGAGCTCAAGCTTTTA 1816
Db 1 TTCTCGAGAAAGTTTGAAGTTTATGATGACATTCGACTTGTGCTCAAGCTTTT 60

QY 1817 TTGGTAAGAACCCACACCTTTTACAGAGTCTGATTAAGCAATTCAGCAACTTCAGCAGG 1876
Db 61 TAAACAGAAACCTCATCTCATAGAGATGCTGATTAACCTCTTCAGCAGCTTCAGCAGA 120

QY 1877 CTAAAGTAATGGCTATGAGATTCCTCGCATGTTGTATGATACACGGAATTAATTTGGGAGA 1936
Db 121 CTAAAGTAACACTCTTGAGAAACAGTGTCTCTCTATGCTCTGCGGAAACCGTAGA 180

QY 1937 TAGACTCGGCTTGAAGAGGAGCTCTGTGCACTGTTATAGCAAAAGTATGATGATGCC 1996
Db 181 TAGACTTCTTGAAGAGGAGTCTCTGTTCATTTGCTCGAGGAGACTTGATGATGCC 240

QY 1997 GTATGTGTTGGGCTTGAAGAGGAGATTCACAAATTAAGAAATCAGCTATTTGGAGT 2056
Db 241 GTTCGTGTTGGGCTTGAAGAGGAGTATGCTCCCTTAAAGAAATACATTAATTTAGAAAT 300

QY 2057 TTGTTTGGAGAAATCAATTCGTATGACAAATATATTCCTCCGACTATGCAAAATTTG 2116
Db 301 TTGTTTGGAGAAATCAATTCGTATGACAAATATATTCCTCCGACTATGCAAAATTTG 360

QY 2117 TGGAAACCTGAGTGGAGGAGGTTGCTTCTCTAGAGTTCAGAGACCAAAAGATTAATAAT 2176
Db 361 TGGAGGATGAGTGGAGGAGGTTGCTTCTCTAGAGTTCAGAGACCAAAAGATTAATAAT 420

QY 2177 TTAACCTCGGAGACTATGATGATCTTATGTTTGAAGTACTTGGAAAGATGGAGG 2236
Db 421 TCAAGCTTGAAGATTAATGATGATCTTATGTTTGAAGTACTTGGAAAGATGGAGG 480

QY 2237 TAAGTCAAGGTTCTCTTGAAGTCTGCTGCAACTATGCAAGATTTGAGCCGAG---- 2292
Db 481 GACAGAGTGGTTCACCTTGGCTGAGCCGAGCTTAAGAGAGATGAGTGGAGAGGCTTA 540

QY 2293 -----CATGTAAAGCTAGTGTATGAGGCACTGCAAAAGTTTCTTCT 2338
Db 541 CTGCAAGTCTTGAATCAATTAAGGCTAGTGAATTCAGGCAATTCAGCAAGGTTTCTTC 600

QY 2339 CCGCTATACAGA 2351
Db 601 TTCGTGCTCAGA 613

RESULT 14
CR486014/c
LOCUS
DEFINITION
Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago
ACCESSION
CR486014
VERSION
CR486014.1 GI:48647590

KEYWORDS
SOURCE
ORGANISM
GSS.
Medicago truncatula (barrel medic)

Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliales;
Medicago.

REFERENCE
AUTHORS
TITLE
JOURNAL
Genoscope.
Submitted (10-JUN-2004) Genoscope - Centre National de Sequencage ;
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrel@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

FEATURES
source
Location/Qualifiers
1. .751

/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="Jemalong A17"
/db_xref="taxon:3880"
/clone_lib="MTH2"
/note="vector: pBelOBAcl1 ; Site_1: HindIII ; Site_2: HindIII ; Cook, D.R. and Kim, D.J.-Genoscope sequence ID : mth2-155019FM2"

ORIGIN

Query Match 7.2%; Score 265.4; DB 9; Length 751;

Best Local Similarity 64.1%; Pred. No. 5.4e-57;

Matches 476; Conservative 0; Mismatches 216; Indels 51; Gaps 3;

QY 1491 ATGATTAACGTTGGGAAAGCTAAATGTTTAAGCCGTGCGGAAATTTTGTGCTG 1550
Db 744 ATGAACATCGAGGCGGAGAGAAAGGTTCTCAAGGTGCTCCCAATTTCTGTGGCAG 665

QY 1551 TTGAGAGAGTGGAGCATCAGCTTGTGGGGGTTGACCCGTGAGAAAGTTATGATG 1610
Db 684 TTGAGAGTGGAGAGAGAGAGCAATTTGCTGGAGTTTACACGTGAGATTCATGATG 625

QY 1611 AGGCGTTTTCAGAAATGACAGCTGTGAGAGGTAATACGTTTATGATA-----CCTTTT 1664
Db 624 AGGCAATTCGCAATGAAGAAAGTGCAGAACAGGTGTTCAATGAATTAATGATGATTT 565

QY 1665 TTTAATTTCTTATGATATTAATTAATTTAGTTTCTCATTTTAATGAT----- 1713
Db 564 TTTAATTTCTTATGATATTAATTTAGTTTCTCATTTTAATGATTTTATTAATTTATCT 505

QY 1714 -----GTTGTGTGAGTGTATCTTTTGTAGCTAC 1745
Db 504 CTTTCATTAATTTCTTTTATCTCAATCTTTTATTAAGTGAACCTTTTGTAGCAAC 445

QY 1746 CCCAAGCAATATTCAGAGAGTCAATTTGAAGTTTACGAAGTTGCACTTGTCTGTGGC 1805
Db 444 ACTTACGAATATTCAGAGAGTCAATTTGAAGTCTTATTAAGTGAACCTTTTGTAGCAAC 385

QY 1806 TCAAGCTTTATTTGTAAGAGCCACACCTTTTACAGATGCTGATPAAGCAATTTCCAGCA 1865
Db 384 ACAAGCTTTGTAAGTAAAGCCACATCTTAACCAAGATGCTGATPAATTTATTCATCA 335

QY 1866 ACTTCAGAGGCTTAAGTAAAGGCTAATGAGATTTCCGTGAGTGTGATGATAC----- 1919
Db 324 ACTTCAACAACTTAAGGTAAACAAATTAAGAGATGCTCCCTCTGTTTATTAATCTCCATGGA 265

QY 1920 ACCGAATTAATTTGGAGATGACTTCGCTAGAAAGGGAATCTGTGCACTGCTTAAGG 1979
Db 264 GATGAGAGAGAGAGAGGATTTGATTCATTAAGAAAGGGGCTGTGTGCACTGCTTGTGG 205

QY 1980 CAAAGTTGATGAATGCCGTATGTGTTGGGCTTACAGAGTGAAGATTCACATATATGAA 2039
Db 204 GAGCTTGATCAATGTGATCAATGTTGGGCTCGATAGTGAAGCTCACTTATGAA 145

QY 2040 TCCAGCTATTTGGAATTTGTTTGAAGATTCAAATCGGATGAGCAATGATGATCC 2099
Db 144 CCACTATTAATTAAGACTTATTAATGAAACGCAAAAGTATGAGACAGTATCTTC 85

QY 2100 TGACATATGCAATTTGTTGGAAACCTGGTTGGACAGGGGTTGTCTTCTCTAGCTTACAGAGA 2159
 DB 84 TGACCTCTGTAATAATTTGTTGGAGACATGTTGATGAGAGTGAGGTTTCCCTAGCTTTAGAGA 25
 QY 2160 CACCAAGATTAATAATTTAAAC 2182
 DB 24 TACTTAAGAGCAAACTTTAAAC 2
 RESULT 15
 COL17047 849 bp mRNA linear EST 16-JUN-2004
 LOCUS GR_EB019K01.r GR_Eb Gossypium raimondii cDNA clone GR_EB019K01
 DEFINITION 3', mRNA sequence.
 COL17047
 ACCESSION COL17047.1 GI:48815734
 VERSION COL17047.1 GI:48815734
 KEYWORDS EST.
 SOURCE Gossypium raimondii
 ORGANISM Gossypium raimondii
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
 1 (bases 1 to 849)
 K1m,H., Yu,Y., Kudrna,D., Hatfield,J., Stum,D., Mueller,C.,
 Udall,J.A., Rapp,R.A., Wendel,J.F., Rao,K., Soderlund,C. and
 Wing,R.A.
 Global assembly of Cotton ESTs
 TITLE Unpublished (2004)
 JOURNAL Contact: Rod A. Wing
 COMMENT Arizona Genomics Institute
 The University of Arizona
 Forbes Building Room 303, Tucson, AZ, 85721-0036, USA
 Tel: 520 626 9595
 Fax: 520 621 1259
 Email: http://genome.arizona.edu
 Plate: 019 row: K column: 01.
 FEATURES
 source Location/Qualifiers
 1..849
 /organism="Gossypium raimondii"
 /mol_type="mRNA"
 /db_xref="taxon:29730"
 /clone="GR_EB019K01"
 /tissue_type="floral"
 /dev_stage="3 to +3 DPA"
 /lab_host="MDH10B"
 /clone_id="GR_Eb"
 /note="Vector: pCMV.SPORT-6.1; Site 1: NotI; Site 2:
 EcoRV; library made by Invitrogen with RNA supplied by
 Wendle lab. Directional cloned into NotI-EV. Colonies
 plated/picked by AGI. More glycerol clones held in -80."
 ORIGIN
 Query Match 6.9%; Score 254; DB 7; Length 849;
 Best Local Similarity 58.7%; Pred. No. 5e-54;
 Matches 518; Conservative 0; Mismatches 325; Indels 39; Gaps 3;
 QY 1828 CCACACCTTTACGATGCTGATGAAGCAATTCACGCACTTCAGCAGGCTAAGTAATG 1887
 DB 1 CCTCATCTCAATTAAGATGCTGATTAACCTCTTCCAGCAGCTTCAGCACTAAGTAACA 60
 QY 1888 GCTATGAGAGATTCCTGGAGATGTGATGATACAGGAAATATTTGGAGATAGACTCGGT 1947
 DB 61 ACTTTTGAAGACAGTGTCTCTCTCTATGCTCTGTGCGAAGCCGTGAGATAGACTTTGCT 120
 QY 1948 CTAGAAAGGGGACTCTGTGCACTGCTTATAGCAAAAGTTGATGATGCCGTATGTGTTG 2007
 DB 121 TTGAGAGGGGCTCTGTTCATGCTGCTGGGGAGCTTGATGATGCGCTTCGAGTTG 180
 QY 2008 GGCTTAAGCAGTGAATTCACATATATGGAATCAGCTATTTGGAGTTGTTTGAG 2067
 DB 181 GGCTTAAGCAGTGAATTCCTCCCTTAAGAAATACATCTATTTGTAATTTGTCTTGGA 240
 QY 2068 AATCAATCGTATGACATGATCTCCTCGAATATGATCAAAATTTGTGAAACCTGG 2127

DB 241 AACTCAAGAGATGACGATGACAGAGATCTCCCGGCTTTGCAACTGCTGGAGGCATGG 300
 QY 2128 TTGGCAGGGGTTGCTTCTCTAGGTTGAGAGACACCAAGATTAATAATTTAACTCGGG 2187
 DB 301 CTAAATGAGAGTGTCTTCTCTAGATTTAGAGACCAAAAGATATACATTTCAAGCTTGA 360
 QY 2188 GACTACTATGATGATCTTATGATTTTGAAGTTACTTGAAGAGTGAAGTTCAGGGT 2247
 DB 361 GATTATTATGATGATCTTCTCTAGATTTTGAAGAGCTTGAAGGAGCAGTGTG 420
 QY 2248 TCTCTTTAGCTGCTGCTGCACTATGAGCAAGATTTGAGCCGAGCATGTGAAGCTAGT 2307
 DB 421 TCACCTTGGCTGAGCGCGAGCTATATGATGAGATGAGAGGCTACTGCAAGTCTT 480
 QY 2308 GCTATGACGAGCAGCAGAAAGTTTCTCCGCTATACAGATAGAAACTCGGCTGAA 2367
 DB 481 GAT-----CATGTAAAGGCTAGTCAATTCAG 507
 QY 2368 CCCAAGATGTCAGAGACAGTGTATGATGATCTGTGTTAACTATGAGCCGT 2427
 DB 508 GCATTTGCAAGAGTGTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 567
 QY 2428 GATGTAAGCCTGCTGCTTATTTGCAAGAGCTGTAAGACCTTGTGAACCTTGAAC 2487
 DB 568 GTGATGATGAACAACCTTCTCTGCTGTAAGAGTGAAGACTCTTGAAGAACTGATCAA 627
 QY 2488 AATATATATCAATTCAGAGCTGGGCTCTC---AGAGAGTCCGTGATGAACACTAGTT 2544
 DB 628 GAGATTTCTGCATCTCAGCTGAGGTTCTGGAATTAAGCAGTTTGAAGAGATGATGA 687
 QY 2545 GAAATGTCCTGCTGATATGTTAAAGAGCAGTGTGAAGATCTAGCTGCTGCTG 2604
 DB 688 GAAATAACATTAATGACAAATTAAGAGTCCAGTGTCAAGATCATGTGCTGCTG 747
 QY 2605 GCAATGACTGATTTCACTGTTCAAGCAGAGATTTT---TCTTAAAGCAGCTCATCT 2661
 DB 748 GTATTTGAGATGATGATGATGTTGCTGTAAGTTTATCTGTAATTTAGTTCAATCA 807
 QY 2662 TTTCAAGCAAGATATGTTTCTTATGAAATCTGATGTC 2703
 DB 808 GTGACCGGTAAAGGCAATCAGTCCAGCTATGCAACTGATGTC 849

Search completed: June 10, 2005, 00:53:15
 Job time : 11203 secs

THIS PAGE BLANK (USPTO)